Agata Lewandowska

The effectiveness of data-driven learning techniques in eliminating Polish advanced EFL learners' interference errors

Praca doktorska napisana
na Wydziale Anglistyki
Uniwersytetu im. Adama Mickiewicza
pod kierunkiem
prof. UAM dr hab. Krystyny Droździał-Szelest

Poznań, 2013

Imię i nazwisko Agata Lewandowska .....
Promotor …… prof. UAM dr hab. Krystyna Droździał-Szelest .....

1. Oryginalny tytuł pracy doktorskiej

The effectiveness of data-driven learning techniques in eliminating Polish advanced EFL learners' interference errors .....
.....

2. Tłumaczenie tytułu pracy doktorskiej

a) na język polski (w przypadku prac napisanych w języku obcym)

Skuteczność technik nauczania opartego na danych językowych (DDL) w eliminowaniu błędów interferencyjnych u polskich zaawansowanych uczniów języka angielskiego .....
.....

b) na język angielski (w przypadku prac napisanych w języku innym niż język angielski)

.....
.....
.....

Podpis promotora

.....

Podpis studenta

.....

Miejsce i data

.....

OŚWIADCZENIE

Ja, niżej podpisany/a

.....

oświadczam,

że przedkładaną pracę doktorską

pt. The effectiveness of data-driven learning techniques in eliminating Polish advanced EFL learners' interference errors

napisałam samodzielnie.

Oznacza to, że przy pisaniu pracy, poza niezbędnymi konsultacjami, nie korzystałam z pomocy innych osób, a w szczególności nie zlecałam opracowania rozprawy lub jej istotnych części innym osobom, ani nie odpisywałam tej rozprawy lub jej istotnych części od innych osób.

Jednocześnie przyjmuję do wiadomości, że gdyby powyższe oświadczenie okazało się nieprawdziwe, decyzja o wydaniu mi dyplomu zostanie cofnięta.

....

(miejscowość, data) (czytelny podpis)

Table of contents

List of tables

List of Figures

List of abbreviations

BNC – the British National Corpus

BYU BNC – the Brigham Young University's interface for the BNC

CA – Contrastive Analysis

CAH – Contrastive Analysis Hypothesis

CALL – computer-assisted language learning

CANCODE – the Cambridge and Nottingham Corpus of Discourse in English

CEA – computer-aided error analysis

CECL – the Centre for English Corpus Linguistics

CEFR – the Common European Framework of Reference for Languages

CI – Confidence Interval

CIA – contrastive interlanguage analysis

CIC – the Cambridge International Corpus

CLAWS – the Constituent Likelihood Automatic Word-tagging System

CLI – cross-linguistic influence

CLT – Communicative Language Teaching

COCA – the Corpus of Contemporary American English

DDL – data-driven learning

EA – Error Analysis

EAP – English for academic purposes

EC – error correction

FFI – form-focused instruction

FonF – focus on form

FonFS – focus on forms

ICLE – the International Corpus of Learner English

IL – interlanguage

IQR – interquartile range

KJO – Kolegium Języków Obcych (*the Teacher Training College*)

KWIC – key word in context

LCIE – the Limerick Corpus of Irish English

LDOCE5 - Longman dictionary of contemporary English (5th edition)

LIBEL – the Limerick-Belfast Corpus of Academic Spoken English

LOB - the Lancaster-Oslo/Bergen Corpus of British English

M – mean

m – median

MED2 - Macmillan English dictionary for advanced learners

MEP – the Multiple Effects Principle

MI – Mutual Information

NKJP – Narodowy Korpus Języka Polskiego (*the Polish National Corpus*)

NLP – natural language processing

NNS – non-native speaker

NS – native speaker

OC – obligatory context

PELCRA – Polish and English Language Corpora for Research and Applications

PICLE – the Polish component of the International Corpus of Learner English

POS – parts of speech

PPP – Presentation – Practice – Production

SLA – second language acquisition

UG – universal grammar

Introduction

The initial idea for this study arose as an attempt to solve problems which recurred in advanced English grammar classes year after year. Students kept making the same mistakes, especially in grammar and word choice, and despite receiving corrective feedback and explicit instruction focusing on the issues which caused the most frequent problems, little progress seemed to be taking place. The errors that were most persistent seemed to be those whose origin could be traced to the students' L1, i.e. Polish. They were not very surprising in tasks involving translation, but they kept appearing in students' spontaneous production as well, both in speech and in writing. Then corpora became available online and at no cost, so it seemed that finally there was a tool that could make a difference, that would assist students in overcoming their problems and in becoming less dependent on their mother tongue, especially in their written work produced outside the classroom. There was an expectation that once students became aware of corpora and were instructed how to use them, they would be eager to refer to them whenever in doubt. It was believed that using corpora for reference would gradually lead to improvement in their spontaneous production as well. It was anticipated that enhanced input in the form of concordances would provide much needed exposure to accurate, native speaker language in areas where learners sought assistance and consulted the corpus to dispel their doubts. That expectation turned out to be overly optimistic. Being introduced to corpora on computer left many students disinterested or, in some cases, overwhelmed with complex query syntax or the number of search options available. Although every year some students became regular corpus users and enthusiasts, most were not interested, and would not consult a corpus unless forced.

The next step, then, was bringing corpus data into the classroom, in most cases in the form of a concordance printout. Before this was done, however, it was necessary to choose what language problems to address with this new, promising tool. To make sure that class time and the preparation process were not wasted on items that were not worth it, a diagnosis of students' problems was required. First a traditional 'manual' error analysis was performed on the participants' examination essays. The material obtained there, however, was limited to the very formal style expected in this kind of writing, and did not constitute a true representation of students' day-to-day use of the target language, which was much less controlled and less accurate. That is where blogs came into play: each year students of English at KJO took a course in information technology, which emphasized the role of Web 2.0 in foreign language education and engaged students in various online projects. One of the key requirements of the course was to maintain a blog in English, in which students not only responded to assigned tasks but also shared their various interests and opinions. Over a few years a large body of student-generated text had been accumulated, which was relatively easy to convert into a corpus of less formal learner English. Thus an opportunity arose to find those aspects of English which are problematic both in strictly formal essay writing and in the slightly more relaxed style of blog postings.

As mentioned above, at an early stage of the study a decision was made to limit it to L1 interference errors. This narrowed the scope of data analysis and gave it a sharper focus. It also opened the study to the use of a parallel corpus, which seemed a particularly interesting instrument to use with a monolingual group of advanced learners. The Polish-English corpus used in the study (Hrdina et al. 2012) holds mostly literary prose and some technical and academic documents, so the level of formality of the texts varied from casual narrative to very formal styles, which approximated the styles of students' work under analysis. The focus on L1-induced errors made the project more challenging, too, because many of the errors selected for the treatment were so persistent that they may have qualified for being diagnosed as 'stabilized' or even 'fossilized'. This is no surprise, as, according to Selinker and Lakshamanam's (1992) Multiple Effects Principle (MEP), transfer is a "privileged", or even "necessary", factor in fossilization. What is more, "[a]pparently fossilized structures will not become open to destabilization through consciousness raising strategies when multiple effects apply" (Selinker and Lakshamanam 1992: 199). One of those multiple effects, according to MEP, is

transfer. The present research project was designed to establish whether the effectiveness of corpus-based teaching in eliminating interlingual errors is higher than conventional modes of teaching. The unique qualities of DDL instruments for classroom work on accuracy and appropriateness, such as access to authentic speech and writing, and the distinctive visual presentation of language data (enhanced input), gave reason to believe that their effectiveness would be higher, too. Whether this is the case is the key subject of this thesis.

The theoretical background for the study includes a wide range of topics. The first chapter begins with an overview of major issues related to the concept of language error. First, readers are presented with a summary of various theories of language learning and their positions on the concept of language error, from behaviorism through the communicative approach and the cognitive approach, to computer-aided error analysis. Then the focus moves towards cross-linguistic influence and interference errors, so as to delineate the field of enquiry more precisely. Attempts are made to clarify the notion of transfer and the ways in which it is thought to affect language acquisition. The chapter ends with a review of various error taxonomies, so as to provide an appropriate framework for the experimental part of the dissertation.

The second chapter begins with a discussion of corpus linguistics and its instruments, followed by an overview of the theories of concept formation and grammar that are related to the field. Next, DDL is set within the context of pedagogy: some general learning theories and language learning theories are presented, to demonstrate their relevance to DDL. The aim here is to achieve a better understanding of the mental processes that learners undergo as a result of corpus-based activities. First, the connection most commonly declared by DDL theorists – cognitive constructivism – is discussed. Then, an alternative view is considered as a possible explanation of the learning processes involved in corpus-based instruction: connectionism and its emphasis on pattern recognition. Further, DDL is viewed from the point of view of those *language* learning theories which seem to share with it at least some of their key notions and ideas. The least 'obvious' one here is Krashen's Natural Approach; it is included in this discussion mainly because of the high importance Krashen attached to providing learners with samples of language and appropriate input. Clearly, DDL cannot be associated with the Natural Approach in a direct way, but it is important to realize that there are some elements that the two approaches have in common: the dependence on authentic language

input well matched to the learner's needs, and the emphasis on perceptive skills rather than production. The other SLA schools of thought that could be linked with DDL and are included in the discussion are the Lexical Approach and Form-Focused Instruction.

In Chapter Three data-driven learning is discussed in more detail. The foundations of DDL were laid by Tim Johns (1991), and then he and his many followers developed them into a consistent language teaching approach (e.g. Johns 1994 and 2002, Granger 1998b, Partington 1998, Cobb 1999a, Hadley 2002, O'Keeffe et al. 2007, and Boulton 2008a). These contributions are reviewed and summarized in the chapter, followed by a catalog of its battery of tasks. The chapter moves on to present an overview of DDL techniques, organized into categories based on the type of source material they engage. The last section of the chapter reviews prior research on the effectiveness of data-driven learning techniques, with the intention of placing the current study in the context of its predecessors.

In the last chapter the reader will find a report on the experimental study and other elements of the research project designed for the purpose of this thesis. In the initial part of the chapter the research questions are stated, and a short characteristic of the participants is provided. Next, an outline of the project is given, followed by a short review of the key research instruments used in the study. The report itself begins with error analysis, first traditional – with students' examination essays – and then corpus-based, with a learner corpus built out of students' blog posts. The analysis is merely an introduction, and its results are then used in the key element of the project – an experimental study on the effectiveness of DDL instruction, carried out in the pre-test / post-test mode. Further, results and statistics for each of the language items addressed in experimental lessons are discussed. The lessons had two versions: the DDL one and the conventional one, and their effects were compared in a statistical analysis, first item by item and then in total.

Apart from the experiment, the project included a survey of students' opinions on the experimental and traditional lessons, with special focus on various techniques and activities they engaged in. A detailed discussion of this survey constitutes the next section of Chapter Four. Students' answers are analyzed and discussed question by question, and provide useful feedback not only in responses to Likert scale questions but also in interesting comments made in response to open-ended questions. The last

section of the chapter contains a correlational analysis in which an attempt is made to find links between the results of the experiment and the outcome of the survey.

At the end a list of references is included, followed by appendices, where the reader can find additional information concerning the error analysis, learner corpus concordances, the test and test results, lesson reports, materials for the experimental and control lessons, as well as the survey data.

Chapter 1: Interference errors in advanced learners' language production

Introduction

Interference errors are usually associated with early stages of foreign language learning, when resorting to L1 (or other languages the learner knows) is a way of coping with the task. Many studies confirm, however, that advanced learners do struggle with interference from their mother tongue as well (cf. Kellerman 1984). What is more, research on error gravity (e.g. Davies 1983) shows that the errors perceived as the most serious ones result from negative transfer[1], because these errors commonly lead to confusion and misunderstanding when learners communicate with native speakers of the target language. This chapter presents the concept of language error, an overview of the ways in which errors were accounted for by various theories of language learning, the phenomenon of transfer itself, and its role in shaping different stages of the development of interlanguage. Teachers' and learners' beliefs regarding corrective feedback are also discussed. The chapter ends with a review of error classification schemes implemented in interlanguage research.

[1] The terms *negative transfer* and *interference* are used interchangeably throughout the text.

1.1. The notion of language error

There are three major options in defining what a language error is. The criteria involved can be correctness (understood as conformity with prescriptive rules of usage), acceptability or grammaticality.

The first of the options – choosing *correctness* as the key criterion in identifying errors – is an arbitrary solution, based on judgments characterized by James (1998: 74) as "recourse to prescriptive normative standards". Such a strict approach, represented by publications like Strunk and White's style guide ([1918] 2009), although still quite common in ELT, does not take into account stylistic features of language, unique to register and genre. One of the aims of this study, however, was to find those areas of English which prove challenging to advanced learners *regardless* of the level of formality. Assuming correctness as the key criterion would be too limiting for this aim to be achieved, and would not be very productive either. The use of language which is in line with prescriptive recommendations might be perceived as inappropriate in less formal contexts. A good example of such differences is the use of *like* as a conjunction in adverbial clauses of manner (e.g. *like they do*). The form recommended by prescriptivists, would be *as*, but *like* is very common in such contexts, especially in spoken language and in informal contexts. The BNC seems to confirm this: a search for "like.[cs*] [pnp*] [do]"[2], i.e. *like* as a conjunction followed by a personal pronoun and the lemma *do*, yielded 222 results, 104 of which occurred in the spoken section of the corpus (and only one in the academic sub-section).[3] Here the 'norm' would reject these sentences as incorrect, but corpus reference proves that such forms do occur in similar, if not identical, contexts, and would hardly be perceived as foreign.

The second option, *acceptability*, is a rather complex concept, characterized by James (1998: 66) as "a practical notion, being determined by the use or usability of the form in question". It is 'context dependent', in the sense that not only linguistic but also non-linguistic factors (e.g. cultural or social) may render a given form acceptable or unacceptable. According to Lyons (1968: 137) "[a]n acceptable utterance is one that has

[2] The query syntax for the BNC interface used in this study is available at the following URL: http://corpus2.byu.edu/bnc/help/syntax_e.asp

[3] The BNC statistics for *like* as a conjunction (per one million words in different sub-sections) are as follows: Spoken: 109.9, Fiction 50.2, Magazine 25.9, Newspaper 13.3, Non-academic 4.5, Academic 2.5, Miscellaneous 5.7 (Davies 2004).

been, or might be, produced by a native speaker in some appropriate context and is, or would be, accepted by other native speakers as belonging to the language in question". Acceptability, then, is closely connected with Lennon's notion of error as "a linguistic form or combination of forms which, in the same context and under similar conditions of production, would, in all likelihood, not be produced by the speakers' native speaker counterparts" (Lennon 1991: 182).

The third criterion that can be applied in error judgments, apart from *correctness* and *acceptability* discussed above, is *grammaticality*. An utterance or a sentence will be judged as grammatical if the structures and forms applied in it belong to the target language system. As Lyons (1968: 137) goes on to say, "It is part of the linguist's task, though not the whole of it, to specify as simply as possible for the language he is describing what sentences are acceptable, and to do this in terms of some general theory of language structure". Acceptability therefore depends on the situational context, as the first part of Lyons's definition says, but at the same time entails grammaticality. For an utterance to be acceptable, it must be congruent with the context in which it is produced, i.e. it must be socially and culturally possible (Lyons 1995: 132), and at the same time must follow the rules of grammar, which operate without context, in absolute and abstract terms. The distinction between acceptability and grammaticality is best explained in the following passage:

> A judgment about the grammaticality of a structure with English words makes a claim about whether the structure "is English" or "is not English," but the judgment need not make any claim about whether the sentence "is good English" or not. Acceptability judgments, on the other hand, make claims such as whether a structure is standard or nonstandard, whether it is easily understood or not, and whether it is stylistically appropriate or not (Odlin 1994: 273).

The discussion above should be enough to demonstrate that grammaticality is a necessary but not a sufficient condition for acceptability. Grammaticality is a component of acceptability, but on its own is not enough for a given form to be accepted as an advanced learner's linguistic choice. At that level, the aim is not only to "get your message across" but also to make the communication as comfortable and effortless on the part of the hearer/reader as possible. That is why the question that should be asked when deciding whether an advanced learner's particular language sequence contains an error should not be "*Can* this be said?" The question should be "Is that what

native speakers *do* say in such contexts/situations?" Admittedly, the approach is not without its problems, because there is no such person as an average or typical native speaker. Every language is a rich and complex entity, with speakers who vary in terms of age, background, social status, geographical location, etc. Each of these factors has an effect on how people use language. The trouble is which variety to choose as the point of reference for the learner, especially in the case of English, now a global language, whose native speakers are actually outnumbered by speakers of English as a second or foreign language. Unfortunately, "[a] totally uniform, regionally neutral and unarguably prestigious variety does not yet exist worldwide" (Crystal 1995: 111).

Still, using acceptability as a criterion for error recognition does constitute a relatively dependable research instrument, especially if it is applied with the support of corpus data. A corpus can be seen as a representation of that idealized native speaker to whom Lyons (1968) and Lennon (1991) refer. That is why error recognition judgments in this study will be primarily corpus-based. Not being a native speaker of English, the author will depend on corpus findings whenever the classification of a particular form or phrase as an error causes any doubts. The corpus selected as reference for this study is the British National Corpus, used for the same purpose in numerous studies and EFL publications, most importantly perhaps the *Longman Dictionary of Common Errors* (Turton and Heaton 1996). Sometimes reference is needed for more recent additions to the English lexicon, especially where American English is the target variety, or when modern technology or media are mentioned. In such cases COCA is used, as it is a monitor corpus updated regularly to include the most recent phenomena in English. The assumption is that if a form features in the corpus with very low frequency or is not represented at all in the BNC or COCA, which are among the biggest, most renowned and dependable corpora available to the general public, then it must be idiosyncratic and is not to be recommended to advanced learners as an available choice. The problem about using Lennon's definition in conjunction with a native speaker corpus in the present project (as corroboration of NNS judgments) is that the students' "counterparts" – referred to in the definition – will not actually be identified, but the general native-English population will have to be depended upon. The BNC is a balanced, representative corpus, in which texts and utterances are organized by text-type and not by language user characteristics. In order to compensate for this, the study will depend on queries within relevant sections of the corpora. Students' examination essays will mainly be measured

against data from the *written* sections of the corpus (academic and non-academic) wherever any stylistic judgments are to be made. The other type of text which is going to be analyzed is the blog post. This genre is not represented in the BNC (or COCA) in a separate section, so here acceptability judgments will be made according to more relaxed criteria, based on the spoken and the less formal written sub-sections of the BNC labeled as "fiction" or "miscellaneous". This is because the language of blogs is generally on the less formal side of the stylistic spectrum. Students participating in the blog project were instructed to use standard English and were aware of the educational context of the task. Still, it is natural that the style of their blog posts *is* informal and relaxed, because that is exactly what their English-speaking peers' blogs are like. Fortunately, with very few exceptions, the students refrained from using slang, offensive language or other objectionable language forms.

The way in which errors are defined is crucial in grammar and usage reference materials, among which Swan's *Practical English Usage* is probably the most popular. His approach is as follows:

> If we say a form is incorrect, we can mean two different things. We may be referring to a form like **I have seen her yesterday*, which normally only occurs in the English of foreigners; or we may be talking about a form like *ain't*, which is used in speech by many British and American people, but which does not occur in the standard dialects and is not usually written. In this book, I am mainly concerned with the first sort of "incorrectness", (...) but I have mentioned a few examples of the second type (Swan 1995: xii).

Defining an error for the purposes of language pedagogy is often based on grammaticality, but with the growing influence of corpus linguistics and *real* language input, the balance is shifting towards acceptability, especially at more advanced levels of proficiency. Swan (1995) takes into account both correctness and acceptability, though not in equal measure. Surprisingly, in the newer edition of the book (Swan 2005) the approach remains unchanged, even though the author explicitly declares that all his explanations and examples were verified against corpus data. Carter and McCarthy (2006), on the other hand, declare acceptability as their key criterion, with distinctions between written and spoken English, regional and standard varieties, and major support from the Cambridge International Corpus.

Various studies in error analysis also need to cope with the intangible nature of many errors. As Lennon (1991) admits, the concept of the language error is very diffi-

cult to define unambiguously, and error identification is a task which is far from straightforward. He quotes a study by Hughes and Lascartou (1982), where 30 experts (ten NNS EFL teachers and English NSs – ten teachers and ten non-teachers) are asked to judge 36 sentences as either correct or erroneous. The inconsistency of these judgments is distressing, especially to those teachers who are not native speakers of the language they teach. If native speakers cannot agree on what is or is not an error, how can NNS teachers feel confident making such judgments and evaluating students on their basis? Nevertheless, they need to do so in their day-to-day classroom practice and testing. A study that is in a sense a follow-up to Hughes and Lascartou (1982) is Hyland and Anan (2006), where more differences in the approach to errors between NSs and NNSs are examined. According to the study, while the former appear stricter and tend to focus on grammaticality and rule-violation, for the latter the more prominent aspects are intelligibility, appropriateness, and sensitivity to style and genre.

Error identification being so elusive, obtaining a set of objective and uncontroversial criteria for analysis is a real challenge. It may be fair to admit after Ellis and Barkhuizen (2005: 56) that there is no 'unproblematic' definition of error – every choice has its weaknesses and raises some reservations. The present author may only attempt to increase dependability by adopting clear criteria for acceptability judgments (as defined by Odlin 1994), a working definition of the concept of a language error in general (as given above after Lennon 1991), and interference error in particular (see section 1.5. , p. 76). Applying these consistently in the error identification procedure, together with the BNC and COCA corpora as references, should be sufficient to make error judgments accurate and consistent.

1.2. Corrective feedback in the classroom

Errors play a crucial role in language learning, whether it takes place in the classroom or in a natural context. In formal education, teachers need to make numerous decisions about how to react to a learner's error, and whether to react at all. If the teacher decides not to ignore the error altogether, s/he has a wide range of choices regarding the form of feedback, depending on the type of the learner's linguistic output, the gravity of the error, rapport with the class, the teacher's convictions or the student's expectations.

The last element is particularly important, as it connects with motivation and other affective aspects of the learning process. Pawlak (2012) reviews a number of studies on students' and teachers' beliefs concerning error correction, and concludes that the results are consistently positive: "[E]ven a cursory look at the available empirical evidence demonstrates that both learners and teachers assume that formal instruction including error correction is necessary and they are convinced of its value, with the former often manifesting much more positive beliefs in this respect than the latter" (Pawlak 2012: 109). Also, learners have shown willingness to receive more feedback on their errors than they do in their lessons, not only about their written work but also about their performance in communicative activities. They indicated their preference for immediate correction rather than delayed feedback, which was corroborated with higher achievement in those learners who did receive such feedback (Griffiths and Chunhong 2008). Interestingly, many participants of that study expressed reservations concerning peer correction and self-correction, indicating lack of confidence as the main cause of concern. One might raise objections that the outcome of the study could have been affected by the unique characteristics of the population on which it was based (Chinese students of English), but a similar study was carried out in Poland, and yielded very similar results (Pawlak 2010, as cited in Pawlak 2012: 111). It is therefore important to keep in mind that, unlike many theoreticians of language acquisition (especially those favoring CLT), people engaged in the actual teaching and learning find feedback on errors indispensable and highly beneficial.

1.3. A historical overview of the concept of error in applied linguistics

The way in which a theory understands the role of error in second/foreign language learning is one of its defining features and reflects the ways in which it conceive of the process of language learning in general. The major theories of language learning will be characterized now in terms of their position on the role of error making in language learning.

1.3.1. The behaviorist view and Contrastive Analysis

Historically, linguists' interest in errors developed within the field known as Error Analysis (EA) in reaction to Contrastive Analysis Hypothesis (CAH), presented by Lado (1957) in his influential publication *Linguistics across cultures*. Contrastive analysis was based on the behaviorist view of language proposed by Skinner (1957), and claimed that a learner's habits and structures developed in L1 might inhibit his/her acquisition of L2, especially regarding those features where the two languages differ significantly. This negative effect of L1 was referred to as *interference*. CAH led linguists to engage in analyzing languages in search of the similarities and differences between them, with the view of making language teaching more effective. Even earlier, in the 1940s, similar ideas were brought into linguistics by Fries, Lado's predecessor as director of the English Language Institute at the University of Michigan, who demanded "an adequate descriptive analysis of both the language to be studied and the native language of the student" (Fries 1945: 5).

From this perspective, errors were proof of the learner's imperfect command of L2, resulting from L1 interference and the "inadequacy of our teaching techniques" (Corder [1967] 1981: 7). What is more, behaviorists believed that it was crucial that errors be avoided, as their recurrence might lead to those incorrect forms being fossilized. Success in teaching was to be achieved by resorting to contrastive analysis (CA), whose task was to predict areas of difficulty for the learner by systematically comparing his/her mother tongue and the target language, and identifying differences between them. Although contrastive analysis was not developed as a new teaching method, but as a framework for language description, Lado (1957) anticipated its application in language teaching, especially in materials design, curriculum development and assessment. It was assumed that differences between L1 and L2 were the main (if not the only) source of errors, and that once those errors were predicted by means of CA, interference could be prevented or minimized through proper instruction.

Teaching techniques of the time were also based on the behaviorist principles, developed within the framework of the audio-lingual method. These typically included oral drills and pattern practice such as memorization, repetition, inflection, replacement, or restatement (cf. Larsen-Freeman 1986). The connection between the results of CA and language teaching would be mainly in the selection of the particular language mate-

rial to be taught, so that the potentially difficult areas of L2 would be practiced extensively and errors prevented. If they did occur, they were to be immediately corrected so as to prevent students from developing bad habits. Fries (1945: 9) claimed that "[t]he most efficient materials are those that are based upon a scientific description of the language to be learned, carefully compared with a parallel description of the native language of the learner". The practical consequence of this view was that language teaching materials were to be designed with speakers of a particular L1 in mind, and not universally for all learners worldwide.

Lado's work on contrastive analysis, strongly criticized in the late 1960s for its behaviorist grounding, was subsequently re-evaluated by some influential linguists in the 1990s, most notably by Selinker (1992), who emphasized Lado's (1957) seminal role in Second Language Acquisition research and claims that further achievements within the field would not have been possible without his theoretical claims, controversial as they may now seem. Moreover, Selinker believes Lado's emphasis on empirical research to be one of his most important contributions to SLA studies, even though it was largely overlooked by his critics. More recently, Kramsch (2007) and Swan (2007) pointed out that a lot of criticism addressed to Lado was actually related to the introductory pages of his book, where he makes general statements about the role of L1 in second language learning and identifies L1/L2 *difference* with difficulty" for the learner, while the details of his further deliberations are commonly overlooked. Now, however, Lado is appreciated for his pioneering approach to problems of intercultural communication, an issue which is in the focus of attention for numerous modern linguists. In the 1970s this part of *Linguistics across cultures* was not of much value to researchers, who were interested mainly in structural aspects of language.

1.3.2. The mentalist view

The behaviorist theory of language was gradually abandoned by linguists after Chomsky (1959) radically criticized Skinner's ideas and gained support from many colleagues, especially psycholinguists, who felt that seeing language as a mere set of habits was too reductive and clearly insufficient in accounting for numerous aspects of how humans develop skills of verbal communication. Instead, Chomsky posited the idea of

language as a uniquely human feature, a set of rules, from which all sentences possible in a given language can be generated.

> It is not easy to accept the view that a child is capable of constructing an extremely complex mechanism for generating a set of sentences, some of which he has heard, or that an adult can instantaneously determine whether (and if so, how) a particular item is generated by this mechanism, which has many of the properties of an abstract deductive theory. Yet this appears to be a fair description of the performance of the speaker, listener, and learner. (...) The fact that all normal children acquire essentially comparable grammars of great complexity with remarkable rapidity suggests that human beings are somehow specially designed to do this, with data-handling or "hypothesis-formulating" ability of unknown character and complexity (Chomsky 1959: 57).

In this view of language, errors are crucial in the process of hypothesis testing: on the basis of language data children try to recognize the rules of grammar and test whatever they have recognized the rules to be by producing and perceiving language. They obtain either positive evidence in the form of the presence of a particular feature in the language input, or negative evidence – either direct (overt correction – not very common in L1 acquisition) or indirect (absence of a form in the input). Chomsky's views did not address the problems of second language acquisition directly at that point, but were very influential in further developments in the field of SLA research.

1.3.3. Error Analysis

Some errors which Contrastive Analysis Hypothesis predicted did not actually occur in L2 speakers' language production, while there were many others whose occurrence it failed to account for, as error analysts later proved (e.g. Corder [1967] 1981, Richards 1971). This was the main argument against contrastive analysis, which soon ceased to find interest and support among researchers and lost its influence on applied linguistics for several decades. Its retreat was also caused by its strong association with the behaviorist theory of language, strongly discredited after Chomsky's criticism discussed above won more and more supporters among linguists. Contrastive analysis was in a way replaced by error analysis (EA), which took the opposite approach to learner language: instead of trying to predict errors on the basis of L1/L2 structural differences, it focused on analyzing those errors which did occur, and thus attempted to gain insight

into the process of language acquisition and make it more effective. It may be said that EA evolved as a testing field for hypotheses and predictions formulated by CA.

The principal figure in the realm of error analysis is Stephen Pit Corder, whose papers, published between 1967 and 1980 and later re-issued as a collection, laid the foundations for research in error analysis (Corder 1981). In his early papers Corder ([1967] and [1971] 1981) still refers to the mother tongue habit as the most commonly recognized source of errors, but at the same time offers an alternative outlook, which takes into consideration the cognitive processes involved in language acquisition:

> The other explanation is that language learning is some sort of data-processing and hypothesis forming activity of a cognitive sort. According to this view [the learner's] idiosyncratic sentences are signs of false hypotheses, which, when more data is available and processed, either by direct observation or by statements by the teacher, i.e., corrections and examples, enable the learner to formulate a hypothesis more in accordance with the facts of the target language. (...) The making of errors is an inevitable and indeed necessary part of the language learning process. The 'correction' of error provides precisely the sort of negative evidence which is necessary to discovery of the correct concept or rule (Corder [1971] 1981: 25).

One of the numerous important assets of Corder's (1981) book is that it proposes a useful algorithm for processing learner language in error analysis (Figure 1). The procedure recommends approaching all learner-generated sentences as potentially idiosyncratic, i.e. ill-formed in terms of L2 rules. This means that every sentence must be analyzed with respect to its conformity with the rules of L2 grammar; Corder uses the word *acceptability* here, but if the terms discussed earlier were to be applied, *grammaticality* would be a more accurate choice. Then, if the judgment is positive, the researcher needs to decide whether the sentence 'makes sense' in its context (*appropriateness*). Only if that is confirmed, can it be accepted as *not* idiosyncratic. The sentence will be recognized as erroneous if either of the two criteria is not met. The procedure is designed in such a way as to prevent *covertly idiosyncratic* sentences from escaping analysis. Such sentences conform to L2 rules by chance, but in fact the rules applied in generating them are part of the learner's *idiosyncratic dialect* (see below) rather than L2. They are superficially well-formed, but they cannot be interpreted successfully in their contexts. The example given by Corder ([1971] 1981: 21) is: *After an hour it was stopped*. On the surface the sentence seems grammatically well-formed; but from the context the reader finds out that reference is made to a *wind*, in which case a passive structure is not possible in natural circumstances. Hence for the sentence to conform to the rules of standard

English grammar it would need to be reformulated thus: *After an hour it stopped*. Such errors are the most difficult ones to recognize, because they cannot be detected by means of superficial scrutiny. A few errors like this were found in the material accumulated for the present study, which is indicated in their descriptions. Errors which were not marked as covert are all overt, which means they are in some ways superficially idiosyncratic.

The next step to be taken after the sentence has been judged as idiosyncratic is to look for a plausible interpretation in the context provided. If this can be done, the sentence is to be reconstructed and comparisons are to be made in order to identify the differences between L2 rules and the rules of the idiosyncratic dialect which generated the sentence. It must be stressed that at this stage the learner's L1 is not involved. Errors are to be identified and interpreted on the basis of a reconstruction of the rules the learner must have formulated. If no plausible interpretation can be arrived at, the researcher needs to resort to the learner's L1. The ill-formed sentence is to be translated literally into L1, and then again a plausible interpretation is sought. If it can be arrived at, the L1 sentence is to be translated back into L2, so that a reconstructed sentence may be provided.

The procedure described above has an important asset: it makes reference to the learner's L1. Those errors which are identified by recourse to L1 literal translation and then retranslation into the TL may be assumed to be interference errors. This very procedure was found highly relevant to the needs of the research conducted for the current study and was frequently applied in building the interference error database for analysis.[4]

[4] It must be said that not all transfer errors can be recognized in this way. For example, those which result from lack of an L2 feature in learners' L1 would not be identified through this process.

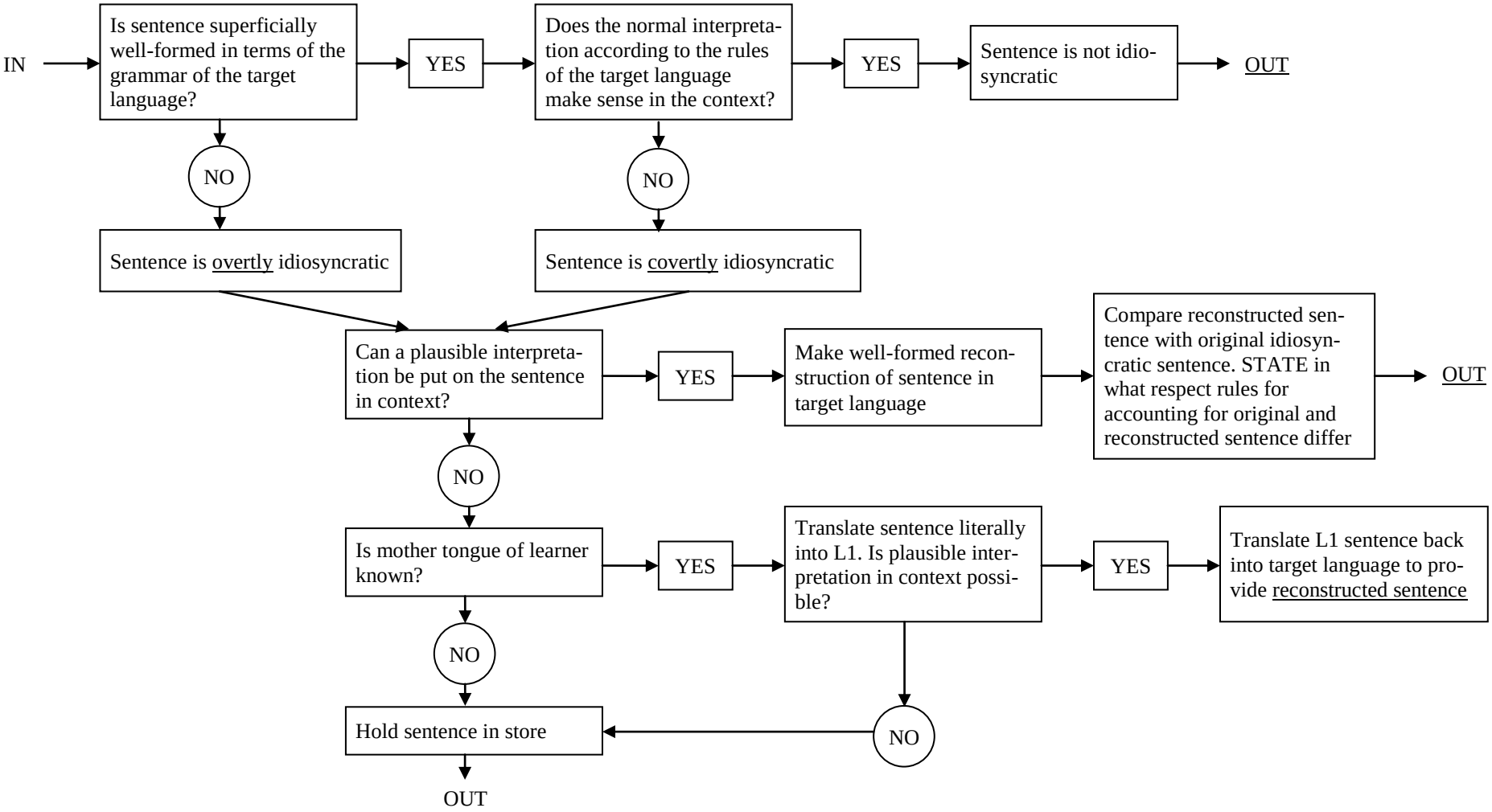


Figure 1. Algorithm for providing data for description of idiosyncratic dialects (after Corder [1971] 1981: 23)

As Lennon (1991) and Ellis (2008) report, in his later work Corder ([1974] 1981) defines the procedure for error analysis in slightly different, broader terms. This was subsequently widely accepted in the field of error analysis as the standard course of action. The procedure is divided into the following five stages:

(1) selection of a sample of learner language;
(2) identification of errors;
(3) description of errors;
(4) explanation of errors;
(5) evaluation of errors.

All these stages require a sequence of decisions on the part of the researcher. First, it needs to be decided what body of learner language is to be investigated. The choice needs to be very well considered, because it may seriously affect the outcome of the study: most commonly, research is performed on compositions and other formal text types. Only a small proportion of EA studies is devoted to spoken language; the reasons are mainly practical, the laboriousness of transcribing speech being probably the decisive one. Then there is the process of recognizing errors, described in detail in the algorithm, where a whole series of yes/no decisions need to be made. The algorithm spreads over the two following stages as well – description and explanation. Finally, an evaluation of the errors is usually performed, especially if the study focuses on error gravity. Such judgments usually involve a panel of experts, who make decisions about particular errors and their perception (e.g. Davies 1983).

Another of Corder's contribution to EA and SLA studies in general is the distinction he made between two major types of data elicitation techniques: clinical and experimental (Corder [1976] 1981: 69), both of which have been employed in the present study. The choice of a particular procedure depends on how controlled students' language production is supposed to be. In other words, if the researcher is interested in the use of a particular form and needs to 'force' learners into using it, experimental elicitation will be appropriate. If the purpose is to observe spontaneous use of language, without any special focus, and then make *post-factum* observations, clinical data collection will need to be applied.

According to Corder ([1967] 1981), errors play three major roles in the process of language teaching and learning. Errors can act as:

- an important source of information about the learner's progress for the teacher;

- useful data for SLA researchers;
- means of receiving feedback for learners on their hypotheses of L2 rules.

Corder's contribution to studies of foreign language learners' errors is best summarized in these three points.

1.3.4. The interlanguage view

The term *interlanguage* was coined by Selinker (1972: 214) as "a separate linguistic system based on the observable output which results from a learner's attempted production of a TL norm". There are, however, several similar terms within the SLA research field, each differing slightly from the others in emphasis and depth: *transitional competence*, *Idiosyncratic Dialect*, the *Approximative system* and, ultimately, *Interlanguage*.[5] Chronologically speaking, the first to be formulated was Corder's ([1967] 1981) *transitional competence*, which he briefly defined as: "[the learner's] underlying knowledge of the [target] language to date". Systematic errors, unlike mistakes (or slips of the tongue/pen), have an important role in defining characteristics of that competence: through analyzing them we can reconstruct the rules which the learner has formulated and follows at a particular stage of his/her language learning. The use of the word *transitional* emphasizes the instability of these rules and their repeated revisions. It may also be understood to be a form of transition from L1 to L2. The concept of transitional competence was later replaced in Corder's publications with *Idiosyncratic Dialect*:

> [A language learner's idiosyncratic dialect] is regular, systematic, meaningful, i.e., it has a grammar, and is, in principle, describable in terms of a set of rules, some sub-set of which is a sub-set of the rules of the target social dialect. His dialect is unstable (we hope) and is not, so far as we know, a 'langue' in that its conventions are not shared by a social group (...), and lastly, many of its sentences present problems of interpretation to any native speaker of the target dialect ([1971] 1981: 17).

The above characterization of Idiosyncratic Dialect covers a broad spectrum of linguistic phenomena: grammar and rules, the social aspect of learner-language and its perception by native speakers. It is therefore not without importance that there is no

[5] Capitalization of some of these concepts is used here in accordance with the original form in which they were introduced.

mention or even indication in it of the learner's mother tongue. Learner language is seen here as an independent entity, governed by its own rules and oriented only towards its final goal – the target language. There are important theoretical consequences of this approach: in strict application of the concept of Idiosyncratic Dialect the very idea of a learner's error needs revising. If the learner speaks a dialect which has its own grammar that is built of learner-generated rules, whatever s/he produces conforms to those rules and therefore should not really be recognized as 'erroneous', 'deviant', 'ill-formed' or 'ungrammatical'. Instead Corder ([1971] 1981) proposes to pronounce learners' sentences which depart from L2 rules as *idiosyncratic*.

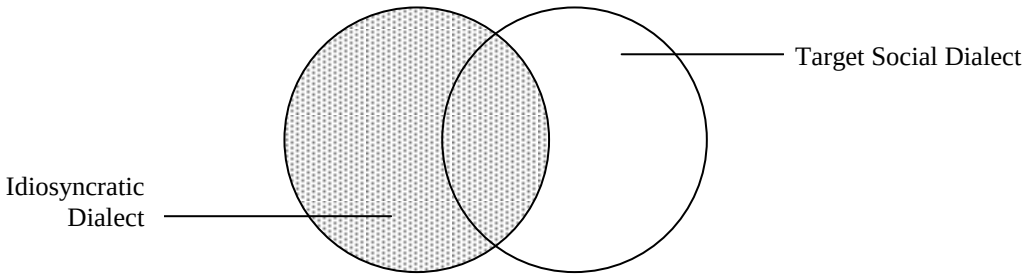


Figure 2. Idiosyncratic Dialect (after Corder [1971] 1981: 15)

Nemser (1971) coined another of the terms referring to learner language, namely: the *approximative system* (L_a). It is a language system that is 'structurally organized' and exists in its own right, independently of L1, which he called the Source Language (L_S), and Target Language (L_T). A series of L_a's come one after another as the learner's command of the Target Language develops, and therefore there is no one L_a, but there are many successive L_a's, which should be indexed as L_{a1}, L_{a2}, etc. This idea, developed within the framework of contrastive analysis, still assigns a very strong role to L1, but stresses the need to analyze the learner's approximative system on the basis of learner-generated data for the sake of developing more successful pedagogical materials and practices.

A slightly different label for a similar notion was adopted by Selinker (1972), who proposes the term *Interlanguage* (IL), which – as the most successful of those discussed here – has become one of the key concepts in SLA research. According to

Selinker (1992: 231), the first definition of interlanguage actually appeared in print earlier, but only as a footnote to a study of language transfer:

> An 'interlanguage' may be linguistically described using as data the observable output resulting from a speaker's attempt to produce a foreign norm, i.e., both his errors and non-errors. It is assumed that such behavior is highly structured. In comprehensive language transfer work, it seems to me that recognition of the existence of an interlanguage cannot be avoided and that it must be dealt with as a system, not as an isolated collection of errors (Selinker 1969, fn 5).

Later, however, the concept of IL was described in more detail, and the following five principal processes of L2 learning were recognized (Selinker 1972: 216-220):

(1) language transfer, by means of which the learner includes L1 features in his/her TL use;
(2) overgeneralization of L2 rules, in which rules of TL are misapplied and the resultant utterances are erroneous;
(3) transfer of training, when a non-TL rule is formed as a result of language instruction in the classroom;
(4) strategies of L2 learning, e.g. simplification, in which the learner reduces the variety of forms available in the TL to those s/he can access;
(5) communication strategies, i.e. building messages in such a way as to communicate one's intended meaning with limited LT resources, e.g. the omission of inflection or function words.

Selinker (1972) uses the term *interlanguage* in two parallel meanings: on the one hand it is a learner's internal linguistic system which operates at a given moment, and on the other a continuum of overlapping 'grammars' which the learner develops over time. Both of these concepts, however, share the idea that elements of both L1 and L2 are part of learner language, which is not the case in Corder's Idiosyncratic Dialect.

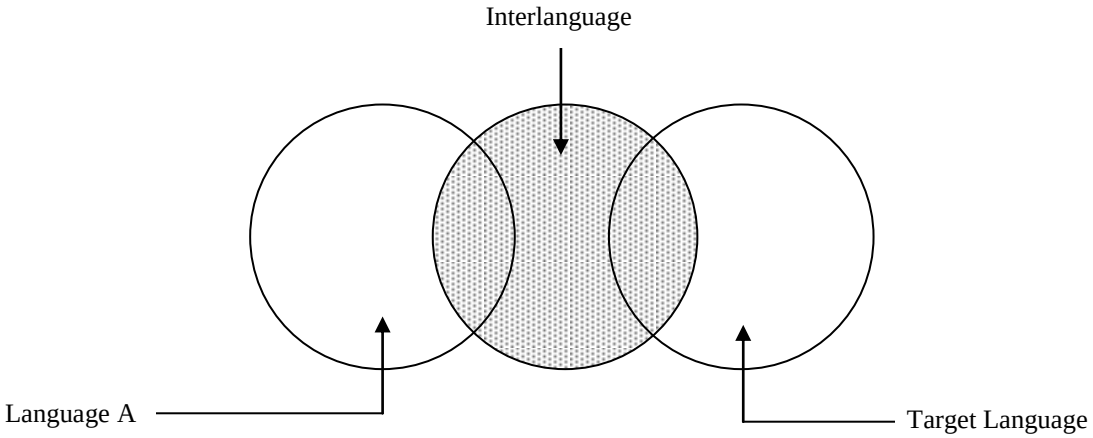


Figure 3. Interlanguage – after Corder ([1971] 1981: 17)

According to Selinker (1972), interlanguage has three main characteristics:

- *IL is permeable*: Learner's rules are not permanent, but are open to correction and change. They gradually evolve.
- *IL is dynamic*: Interlanguage undergoes continuous revision and extension. The learner is constantly engaged in hypothesis testing, which makes his/her language unstable and changing.
- *IL is systematic*: Its development is not haphazard, but follows a predictable order.

Cook (1993) observes that the value of the idea of interlanguage in SLA research lies not only in the processes that Selinker proposed as part of language learning and production, but also in his insistence that these phenomena need explanation in terms of mental processes. The model does not focus merely on the linguistic aspect, but broadens its scope to psycholinguistic considerations, asking how the processes of foreign language learning unfold. This opened a new area within applied linguistics: Interlanguage studies.

As far as the role of errors in the model is concerned, they are strongly connected with the five processes involved in IL listed earlier: they may be an effect of L1 (negative) transfer, if inadequate interlingual identifications are made between L1 and L2; they may also result from overgeneralization, transfer of training, or learning and communication strategies. The repertoire of errors changes constantly in concert with the evolution of the learner's range of grammar rules. More importantly, one of the central features of the Interlanguage Theory is *fossilization:* "Fossilizable linguistic phe-

nomena are linguistic items, rules, and subsystems which speakers of a particular NL will tend to keep in their IL relative to a particular TL, no matter what the age of the learner or amount of explanation and instruction he receives in the TL" (Selinker 1972: 215). Fossilization has two variants: *favorable*, if the form that becomes stabilized is correct, and *unfavorable*, if the form that fossilizes in IL does not follow L2 rules. In this case the form becomes what may be called a permanent error. Another of Selinker's ideas associated with learner errors is *backsliding*, which refers to frequent situations in which the learner uses an erroneous form even though it has been previously 'eradicated' from his/her production performance. Such instances are understood to be caused by special conditions in which the production takes place, such as a high complexity of the message, high level of stress or, quite the opposite, extreme relaxation. Selinker understands the explanation of this phenomenon to be a crucial element of any language acquisition theory.

1.3.5. The communicative view

The broadly understood communicative approach was developed in Great Britain in the early 1970s, when the social aspect of language became more prominent within linguistics. Its proponents, such as Henry Widdowson and Christopher Candlin, built on new work within the sociolinguistics and functional linguistics of the time. The key concept here is Hymes's (1972) *communicative competence*, which evolved in response to Chomsky's very narrow and abstract understanding of competence as the ubconscious knowledge of grammar rules in a given language. For Hymes, communicative competence is a much broader and deeper concept – not only an inherent grammatical competence, but also the ability of language users to convey and interpret meaning in a variety of communicative situations. Hymes emphasized the sociolinguistic aspect of communication, and so communicative competence encompassed the language user's knowledge

1 whether (and to what degree) something is formally *possible*;
2 whether (and to what degree) something is *feasible* in virtue of the means of implementation available;
3 whether (and to what degree) something is *appropriate* (adequate, happy, successful) in relation to a context in which it is used and evaluated;
4 whether (and to what degree) something is in fact done, actually *performed*, and what its doing entails (1972: 281).

These aspects of language knowledge became the focus of attention for communicative language teaching (CLT), thus changing the emphasis from the formal rules of the language system to the social aspects of communication. Hymes's notion of competence was later elaborated upon by Canale and Swain (1980), who divided it into four components: *discourse* competence (the ability to connect sentences or utterances into meaningful texts), *sociolinguistic* competence (the ability to use language adequately to the situation), *strategic* competence (the learner's ability to compensate for gaps in his/her knowledge of the target language), and *grammatical* competence (the learner's knowledge of the elements of the language system).

Another cornerstone of communicative language teaching perspective is the functional approach to language developed by Halliday (1973). One of his numerous contributions to linguistics is a new idea of grammar:

> The basic opposition, in grammars of the second half of the twentieth century, is not that between 'structuralist' and 'generative' as set out in the public debates of the 1960s. There are many variables in the way grammars are written, and any clustering of these is bound to distort the picture; but the more fundamental opposition is between those that are primarily syntagmatic in orientation (by and large the formal grammars, with their roots in logic and philosophy) and those that are primarily paradigmatic (by and large the functional ones, with their roots in rhetoric and ethnography). The former interpret a language as a list of structures, among which, as a distinct second step, regular relationships may be established (hence the introduction of transformations); they tend to emphasize universal features of language, to take grammar (which they call 'syntax') as the foundation of language (hence the grammar is arbitrary) and so to be organized around the sentence. The latter interpret a language as a network of relations, with structures coming in as the realization of these relationships; they tend to emphasize variables among different languages, to take semantics as the foundation (hence grammar is natural), and so to be organized around the text, or discourse (Halliday 1985: xxviii).

This concept of grammar as a system related to meaning and interpersonal relations goes back to Malinowski (1923) and his ethnographic view of language, and is as distant from Chomsky's "armchair linguistics" as can be.

Halliday defined seven language functions (or "models") which a child needs to be able to use in order to communicate successfully with others (Halliday 1973: 11-16). These are:

(1) the instrumental function: using language to get things done;
(2) the regulatory function: using language for exercising control over others;
(3) the interactional function: using language for interaction between self and others;
(4) the personal function: using language as a form of individuality;
(5) the heuristic function: using language to explore the environment and investigate reality;
(6) the imaginative function: using language to create a world of imagination;
(7) the representational function: using language to communicate information and express propositions.

The seven functions are key components of meaning as understood by the Systemic Functional Grammar. Proponents of the communicative approach believe that learning a language is on a par with learning to perform these functions. Finally, it is important to point out that according to Halliday (1985: xiv), in functional linguistics "syntax and vocabulary are part of the same level in the code", and the two together should be seen as one: lexicogrammar, i.e. a set of potential choices that a language user has in order to express his/her intended meaning. This unity of grammar and vocabulary was then strongly emphasized by Sinclair (1991) and became the standard position in corpus studies of language. Also Lewis (1993) saw lexicogrammar as one of the key notions of his Lexical Approach. The same position was adopted in the error analysis performed for the needs of the present study.

According to Richards and Rogers (2001), CLT's practical recommendations concerning language teaching can be summarized in the following major principles, advocated by such authors as Littlewood (1981) and Johnson (1982):

- *the communication principle*: activities that engage learners in real communication facilitate learning;
- *the task principle*: activities in which learners perform meaningful tasks facilitate learning;
- *the meaningfulness principle*: language that is meaningful to the learner enhances the learning process.

Generally, CLT emphasizes practice and meaningful language production as a way of developing fluency and communicative skills. Courses are developed mostly on the basis of notional-functional syllabuses rather than grammatical structures. Finally, learners' needs are an important element of course design.

Attractive as the above principles seem, the communicative approach does have its weaknesses, and these are mostly connected with the way it deals with errors. With so much emphasis on communicating one's meaning, it often leads to highly inaccurate language performance, and with large amounts of classroom discourse being produced, little corrective feedback can be offered by the teacher. The assumption is that feedback will be provided in the actual process of communication, but this assumption is highly questionable. This perhaps would be the case if the language exchanges took place between learners and native speakers of the target language. As it happens, language tasks are often performed within learner groups and even if the other learners do react negatively to one learner's choice of language, they do not have enough authority to make the learner revise his/her choice of language form for a particular intended meaning. The result often is a fossilization of errors, and high fluency at the expense of low accuracy (cf. Meriläinen 2010). Communicative language teaching may be more effective in the context of *second* language learning, where native speakers of the target language are accessible outside the classroom in the learner's natural environment. In the context of *foreign* language learning, however, communicative teaching does not seem to be very effective in developing learners' accuracy. This is why various techniques of form-focused instruction need to be introduced into the foreign language classroom, corpus-based techniques among them.

The position of the communicative approach on learners' errors is clearly formulated by Widdowson (1979), who offers a unique and controversial view of the problem. First of all, he questions Corder's ([1967] 1981) distinction between errors and mistakes, raising the case of "lapses which are persistently regular but which the learner knows are lapses once his attention is drawn to them" (Widdowson 1979:186). On the one hand, they seem to be typical cases of regular, or even fossilized, errors, but on the other, the rules they violate are not absent from the learner's language system. Perhaps they have not been fully internalized yet, but they are part of competence (the learner is able to refer to them, or even self-correct, when alerted to the problem). These situations are especially common with advanced learners.

Also, learners tend to commit certain errors in speech, but not at all, or very seldom, in writing. Such inaccuracies are systematic then, but only in specific conditions of performance. Should they be recognized as errors or mistakes? Widdowson (1979) resolves this by distinguishing between two sets of rules: *context* rules (also referred to as *use*), which describe what the learner actually does, and *code* rules (or *usage*), which represent the learner's conscious knowledge of the system. The two sets are not identical, the latter being used as a reference, but not necessarily controlling actual language production. The discrepancy between what the learner knows and what s/he does is accounted for through an analogy to working class speakers who approximate to standard usage only when focusing their attention on how they speak, but deviate from standard in normal situations of interaction. The argument, however, may be undermined by considering language tests, in which learners are highly focused on their language performance and still cannot prevent themselves from breaking code rules. The very idea of questioning one of the basic distinctions of error analysis is worth emphasizing, however. Computer-aided error analysis – which is currently the dominating type of research on learner error within SLA studies – does not observe the distinction either (see section 1.3.8.).

The key notion employed by Widdowson to account for learner errors is *simplification*, defined as

> the process whereby a language user adjusts his language behaviour in the interests of communicative effectiveness. This adjustment may involve a movement away from the reference norm of the standard language so as to arrive at forms of speaking judged to be dialectically appropriate in certain contexts of use. (...) The adjustment may involve either the increase or decrease in complexity of usage (1979: 189).

It is clear from the above definition that the concept originates from the native use of language, but it is also applicable to foreign language learners. Widdowson stipulates that errors which occur as a result of language transfer, transfer of training, learner strategies and the other processes defined by Selinker (1972),[6] occur because learners try to adjust the language system they are learning to their communicative needs just as native speakers do. "Errors are the result of the learner's attempt to convert linguistic usage into communicative use" (Widdowson 1979: 190). The final conclusion, which deter-

[6] See section 1.3.4. p. 34

mined the general attitude of CLT to learner errors, is that they are actually evidence of success rather than failure, and that the failure to observe particular code rules is a result of success in generating context rules. Widdowson's discussion of these issues presents the language classroom as a highly confrontational situation: the teacher in the role of an 'oppressor' who imposes a rigid formal system, and the learners as creative 'rebels', who despite the authoritarian imposition manage to maintain the communicative nature of language and approximate their communicative effect.

Such strong statements led to classroom practices which were extremely tolerant of errors, and which often resulted in a low accuracy of learners' language production, fluent as it might have been. In cases like this, it might be particularly hard for learners to make progress to the advanced level in their language proficiency, where accuracy is expected and aspired for. Fortunately, not all teachers following the communicative approach assumed such an extreme position. Some research has shown that in communicative lessons of English as a second language taught by well trained teachers to highly motivated learners, the rate of Focus on Form episodes was relatively high, with both reactive and preemptive types of focus. Both are incidental, as all FonF episodes should be, but the former is usually initiated by the teacher (or sometimes by another student) in response to an error that has occurred in the class discourse, whereas the latter can be initiated by either the teacher or the student whenever a language element which has emerged causes doubts or is predicted to be a source of error in the future (Ellis et al. 2001). This may be a short incident, for example, the teacher making sure that the word which has come up is spelled correctly, or a student asking about a choice of a preposition for a particular context *before* any error is actually committed. The study shows that those episodes which were initiated by students were the most effective.

1.3.6. The Natural Approach

A special status within the broadly understood communicative language teaching must be assigned to the Natural Approach shaped by Krashen and Terrell (1983). Its American origins and some differences in theoretical assumptions do not allow the Natural Approach to be fully identified with communicative language teaching. It cannot be overlooked, however, that many of Krashen's ideas do coincide with the communicative

approach. Both emphasize natural communication as a necessary element in language learning, and challenge the effectiveness and justifiability of explicit grammar teaching. The important difference is a much deeper theoretical insight into SLA processes, with special focus on *interlingual errors* and their sources (e.g. Dulay et al. 1982). Briefly, such errors are believed to originate from two main phenomena characteristic of a foreign language classroom: one is the pressure to perform, and the other is poverty and poor quality of L2 input. Both of these are also intrinsic in a typical communicative classroom, where students are encouraged to use L2 as much and as early as possible. The result is that they need to resort to L1 in search of any way to communicate their message, because they are not ready to perform successfully in L2. What is more, they are exposed to large amounts of 'less-than-proficient' L2 production from their colleagues, which adds to the confusion and may be an additional source of errors. Similar observations were made by Ervin-Tripp (1974) in a study of 31 English-speaking children learning French in conditions of total immersion in Geneva:

> Most of the evidence showing mother-tongue interference in the learning of syntax has had two peculiarities: It has come from learning conditions in which the second language was not the language of the learner's larger social milieu so that the learning contexts were aberrant both in function and frequency of structures. Further, both the learning and the testing often occurred in situations where the milieu and the addresses were not overwhelmingly connected with the second language. Yet we know that learners are extremely sensitive to such nuances (Ervin-Tripp 1974: 121).

Further, Ervin-Tripp (1974) links dependence on L1 and other known languages with the semantic complexity of the utterance and the learner's readiness to express it:

> I would suppose that if we push a child to generate sentences about semantically difficult material or concepts unfamiliar in the new culture, he may use somewhat different production patterns. (...) I am suggesting that the simpler the semantic task, and the simpler the relation between meaning and form (e.g., description vs. inference), the less the likelihood the speaker will have recourse to other-tongue formations (1974: 121).

Both Krashen and the proponents of communicative language teaching have a similar, though not identical, stance on how teachers are supposed to react to errors. In CLT errors are believed to be an unavoidable and necessary element of language learning, in the course of which learners are able to arrive at appropriate L2 rules by means of trial and error through an intensive use of the language in communicative activities (Richards and Rogers 2001). Some incidental focus on form is acceptable, though im-

plicit feedback (e.g. recasts) is generally preferred, so that the learner could keep the focus of his/her attention on meaning rather than on form. It must be noted, however, that Lyster and Ranta (1997) found this form of feedback the least effective, when compared with elicitation, metalinguistic feedback, clarification requests, and repetition. The Natural Approach has a *non-interventionist* position on the matter of correcting learner errors. Explicit corrective feedback is viewed as pointless, because conscious knowledge of rules is believed not to translate into the learner's language competence. Only through a sufficient amount of suitable input ($i + 1$) is the learner able to acquire L2 rules and follow them in his/her language production. What is more, according to Krashen and his followers (e.g. Dulay et al. 1982), the processes of second language acquisition are believed to mirror those of first language acquisition, with the Affective Filter Hypothesis accounting for the differences in the outcomes. This is why the input should be as similar to the L1 learning environment as possible. Krashen's point is that children learning their first language are very seldom corrected; instead, adults are observed to focus on the content of the message delivered by the child, showing him/her how successful the communication has been. On the basis of this information the child verifies the language rules s/he has been testing and either accepts or revises them. According to the Natural Approach, classroom language learning should follow the same path.

1.3.7. The cognitive model

A very different position on second/foreign language acquisition was developed by cognitive linguists. Adult language learning is seen to be the same as any other type of learning, in which there is no access to the innate language acquisition system available to children. This basic assumption is referred to as the Fundamental Difference Hypothesis (Bley-Vroman 1989).

Within the cognitive school of thought Gass (1997) developed an integrated model of second language acquisition, which takes into account a sequence of stages leading from input to output. Acquisition is not seen as a final outcome, a state, but is defined here as "a *process* beginning with input apperception and culminating with integration of new linguistic information into an existing linguistic system, output then

being the manifestation of newly integrated or acquired knowledge" (Gass 1997: 4). The trigger for the process presented in Figure 4 is *input*, which is never processed by the learner in its totality, but only some of which is utilized in language acquisition. This is demonstrated by the dotted line, which represents the filter between input and acquisition itself. Gass and Selinker (1994: 298-299) explain what factors prevent all of the input from being processed by the learner. The first one is *frequency*. Elements that are very common in input may be noticed and processed more easily. On the other hand, features that are particularly infrequent may also attract the learner's attention, especially at a more advanced level. The second factor is *affect*, a concept referring to such issues as social distance, status, motivation and attitude, which can all limit the amount of information the learner perceives in the language input. Further, Gass and Selinker (1994) refer to prior knowledge as a limiting factor in input processing: this may be the native language or other languages, it may be the current knowledge of the TL, other background knowledge or language universals. All these have a strong impact on which elements of input and how much of it will be allowed in for further processing. Finally, a factor central in achieving success in learning: *attention*. Without attention there can be no cognitive processing, yet to a large extent this particular factor is difficult for the learner to control, sometimes for trivial reasons like tiredness, health, stress or other personal issues of the learner, sometimes because of the input itself being inadequate in terms of quantity or quality.

The first stage of the acquisition process itself is *apperceived input*, that is, the part of input which has evoked the learner's *apperception*, the recognition "that there is a gap between what the learner already knows and what there is to know," or "the process of understanding by which newly observed qualities of an object are initially related to past experiences" (Gass 1997: 4).

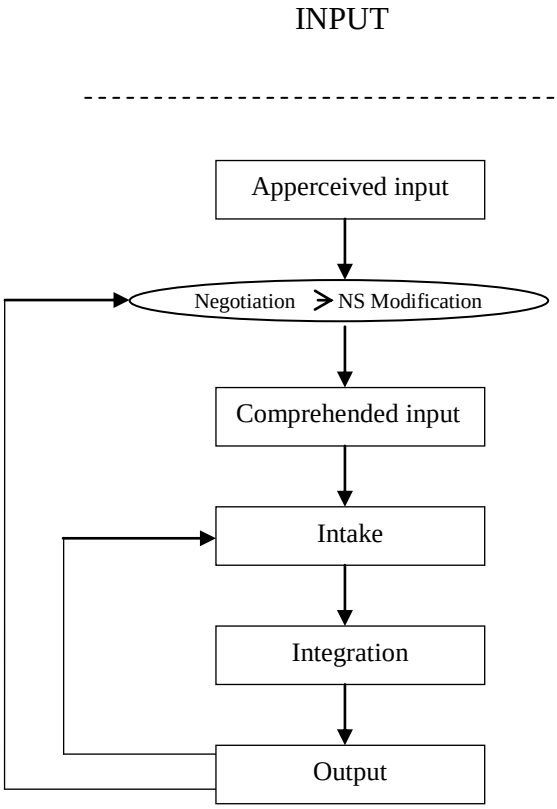


Figure 4. An integrated model of second language acquisition (based on Gass 1997: 3)

This concept is closely associated with Schmidt's noticing hypothesis (1990), which maintains that attention to input is conscious and involves *noticing the gap*, i.e. the learner recognizes the difference between the input that is provided and the output s/he is able to produce at a given moment (see section 2.6.3.). Later, Tomlin and Villa (1994) distinguish three elements of attention: *alertness* (motivation and readiness to learn), *orientation* (the general focus of attention) and *detection* (the cognitive registration of language facts which triggers further processing). The three elements are clearly relevant to remedial teaching and classroom procedures which address learners' errors. Having recognized a recurring error, the teacher can help students understand why a given form should be avoided in particular contexts, for example by demonstrating the confusion it may cause in communication with NSs (alertness). The next step involves making students process the necessary information, most commonly by means of analyzing examples of language use and misuse (orientation). Finally, students will have an opportunity to draw conclusions and to verify their understanding of how a given form is used in TL and what meaning it carries (detection).

According to Gass's model, then, the very first stage of acquisition involves the learner's cognitive, conscious processes, a view quite opposite to what proponents of CLT and especially of the Natural Approach claimed to be the case. The next stage is *comprehended input*, which Gass insisted to be distinguished from *comprehensible* input, a concept coined by Krashen (cf. Dulay et al. 1982). The difference between the two lies mainly in the former being learner-controlled ("the learner is doing the 'work' to understand" – Gass and Selinker 1994: 300), whereas the latter is controlled by the person providing the input (usually the teacher or a NS deciding what language forms the learner is ready to process). Another difference is that comprehensible input is a dichotomous variable: input is either recognized as such or not. Comprehended input, on the other hand, is assumed to be multistaged; that is, there are various degrees of how deeply the learner has been able to understand the meaning – from the level of semantics, where the general meaning of a given word sequence or phrase is grasped, to the deeper syntactic analysis (parsing), where the phrase is broken down into its components, whose functions and features are then recognized. Apperception of input plays the role of "a priming device", and in a way paves the way for comprehension, which, in turn, is affected significantly by prior linguistic knowledge (L1, L2, L3, etc.). "Comprehension cannot take place in a vacuum" (Gass and Selinker 1994: 302), and needs some foundations to build upon.

The amount of input that has been comprehended is not identical with *intake*, another independent component of the acquisition process, because some of the input may be used merely to handle an exchange that the learner has become engaged in 'on the spot', but without any permanent effect on his/her interlanguage. Intake, then, is "the process of assimilating linguistic material. It refers to the mental activity that mediates between input and grammars. (...) Some of the major processes that take place in the intake component are hypothesis formation, hypothesis testing, hypothesis rejection, hypothesis modification and hypothesis confirmation" (Gass and Selinker 1994: 302-303). Hence, it can be said that it is from this stage of acquisition that errors (as distinguished from *mistakes* by Corder [1967] 1981) originate. If the input is miscomprehended, because of its limited amount or poor quality, for example, and the learner's hypotheses are false, then errors occur. What is more, if there is no apperceived and comprehended input to the contrary, the hypothesis is never revised and may remain a permanent element of the learner's metalanguage, the result being fossilization. Accord-

ing to Gass (1997), these processes depend largely on language universals and the learner's L1, both of which constitute the mental framework for them. Importantly, Gass emphasizes that the processes in question are enhanced by L2 interaction.

The next stage of Gass's model is *integration*. There are four different phenomena that can occur at this stage:

- *integration*: A confirmation or rejection of a current hypothesis takes place, which results in a step forward in the development of the interlanguage.
- *apparent nonuse*: There is no difference in the rules of current interlanguage, but rule strengthening occurs, which may have a positive effect on the learner's fluency ("automatic retrieval" is enhanced).
- *storage*: The learner is not ready to incorporate a portion of intake into his/her interlanguage, but the form is stored in the memory and accessed when an opportunity for explanation arises, or subsequent input confronted with the item in storage allows the learner to form, confirm or reject hypotheses.
- *nonuse*: Input is neither incorporated into the learner's grammar nor stored for further use; it may not have been fully comprehended or cannot be linked to the learner's knowledge base in any way.

Gass and Selinker (1994) point out that integration is dynamic and interactive: it may be stretched in time and may change from one form to another. Also, it is bound to other components of acquisition and is not an independent element. The developments are evident in the learner's language output, where rules of his/her grammar can be seen operating. It may be, however, that there is no visible change in the language, and still modifications are made: the rules become re-analyzed and change from very narrow ones (applied to one particular form) to rules used productively to generate many forms of the same type (e.g. derivative suffixes in word formation).

The final component of the model of second language acquisition is *output*, which "represents more than the product of language knowledge; it is an active part of the entire learning process" (Gass and Selinker 1994: 307). Its role is dual: first, output forces the learner to analyze language at the syntactic level, because when generating new language the learner depends on the knowledge base already acquired and uses the rules of his/her grammar in a productive way, as 'building blocks'. This, on the other hand, is presumed to strengthen comprehension in further interaction and enhance comprehended input. Output also allows negotiation of meaning and thus works towards

comprehension. The second role for output is to allow the learner opportunities to receive feedback on his/her hypotheses and thus enhance the intake component. When producing language in spoken interaction or in writing, the learner has an opportunity to test new rules or verify previously formulated ones. These two functions of output are represented in Gass's model in Figure 4 by two arrows, one of which extends from Output (at the bottom of the diagram), to the stage immediately preceding Comprehended Input, and the other connecting with Intake. In this way the model accounts for two channels of feedback in second language acquisition processes.

Output itself is not an identical reflection of the grammar system developed through integration of input. The gap between the two may be seen as parallel to the differences between the native speaker's competence on the one hand, and performance on the other. Many factors affect outcome: the fundamental ones are the rules of the learner's grammar and his/her linguistic knowledge, but there are also other aspects of communication involved, such as the mode (written or spoken), the learner's personality or the type of situation the output is generated in (Gass 1997).

Output, however, is the only language material available to SLA research. There is no direct access to that intangible, abstract entity – the learner's grammar. The important result of this is that in practical terms it is very difficult to distinguish between what Corder ([1967] 1981) defined as *errors* and *mistakes*. The fact that an unacceptable form appears only once in a sample of a learner's output does not guarantee that this is a case of a mistake. The target language item in question may have a very low frequency rate, and therefore is not very likely to recur in a short stretch of text. This is even more probable in advanced learners, whose language repertory includes lower frequency words, phrases and structures, especially in formal contexts. Performance level faults may evidence instability of a given target rule in the learner's interlanguage, which indicates that some further work or more input is needed for its integration to be complete. That is one of the reasons why the error analysis performed in the present study does not distinguish between errors and mistakes, with the exception of such obvious typographical errors as **hte* instead of *the*, for example, which are treated as minor incidents or slips of the pen, and therefore ignored.

1.3.8. Corpus-based error analysis

The development of corpus linguistics in the 1990s brought about new interest in error analysis, owing to the possibilities offered by exploring learner corpora. The new technology and research methodology, defined by Sinclair (1991), allowed researchers to obtain information about features of learner language that was unavailable with the more traditional tools of EA. There are some fundamental theoretical differences between traditional EA and corpus-based EA: the distinction between *competence* and *performance* is not sustained, which makes a significant difference in the analysis:

> The new option opened up by the computer is to evaluate actual instances and select the most typical. A complete set of typical instances should exemplify the dominant structural patterns of the language without recourse to abstraction, or indeed generalization. (...) It is, therefore, unnecessary to make a sharp distinction between abstract and actual language structure – the sort of distinction embodied in Saussure's *langue* and *parole* or Chomsky's *competence* and *performance* (Sinclair 1991: 103).

The logical consequence of this is the dismissal of the key distinction in EA – that of *error* and *mistake*. The only reliable source of information about learners' interlanguage, say corpus analysts, is their language performance, and this is what should be analyzed. Another, more practical, difference is that the amount of analyzed language material can be considerably larger with corpora, usually holding hundreds of thousands of words of learner-generated texts. Traditional EA was usually limited to much smaller collections, for example:

- Dušková (1969) 8,500 words (50 papers of ca 170 words);
- Ghadessy (1980) 10,000 words (100 papers of ca 100 words);
- Rogers (1984) ca 14,500;
- Lennon (1991) 21,000 words (60 recorded narrations).

A large size of a sample allows more accurate judgments and generalizations. The methodology of corpus research, however, also brings many more important innovations to error analysis. Previously the focus of attention was directed at learners' errors only. Now, with corpus-searching facilities, error statistics can be set against quantitative data of the 'correct' use of a given form or pattern, which gives a more accurate picture of learners' interlanguage and of their difficulties. Needless to say, corpus analysis proves especially relevant to analyzing advanced learners' interlanguage. Their

production is more abundant, more varied and therefore more difficult to inspect manually. Lower frequency words and structures – absent or very rare in beginners' or intermediate learners' production – occur more often, and their use can be analyzed through corpus techniques. The scope of such analysis typically includes lexicogrammatical patterns, as these aspects of language yield themselves to corpus analysis most effectively and produce the most interesting results. However, such aspects of learner language as features of discourse and genre have also been investigated with learner corpora (e.g. Granger 1997, Ragan 2001, Upton and Connor 2001).

An academic institution which is probably the most active in the field of learner corpus research is the Centre for English Corpus Linguistics (CECL) at the University of Louvain, with its director Sylviane Granger. Among the center's numerous achievements is the compilation and publication of ICLE, the International Corpus of Learner English (Granger 2002), a collection of argumentative essays[7] written by higher-intermediate to advanced EFL students of 11 different L1 backgrounds. Learner corpora usually represent one type of text, and the most common genre is an argumentative essay (Ellis and Barkhuizen 2005). This is the case with the ICLE project, too. The corpus is widely used as a research tool by linguists interested in interlanguage, as it allows comparisons to be made between various L1 speakers' English output, including errors. This type of research is relevant to one of the most contentious issues in EA: which, and how many, errors are interlingual and which intralingual. It is important to note that there is a Polish component of ICLE, called PICLE, compiled by Kaszubski (2002) at the School of English, Adam Mickiewicz University. These resources were also employed in the study performed for this dissertation.

The CECL employs two learner-corpus-based methodologies: contrastive interlanguage analysis (CIA) and computer-aided error analysis (CEA). The former analyzes learner data in comparison with either native speaker data or interlanguage data from different groups of learners, defined by such factors as L1, level of proficiency, age, etc. The latter involves 'word-by-word' analysis of all the material in a given corpus, which follows the procedure described below:

> First, the learner data is corrected manually by a native speaker of English who also inserts correct forms in the text. Next, the analyst assigns to each error an appropriate error

[7] It is usual for learner corpora to be built of texts of the same type.

> tag (a complete list of all the error tags is documented in the error tagging manual) and inserts the tag in the text file with the correct version (Dagneaux et al. 1998: 165).

The tagging process is very time consuming and requires substantial resources, but its results open new possibilities for error analysis and interlanguage studies. In the thus annotated corpora quantitative information can be obtained and comparisons between various groups of learners can be made concerning underuse, overuse and misuse of particular forms. These aspects of interlanguage were previously unavailable to traditional error analysis. Knowledge of this kind is of high practical use in the realm of ELT.

Another Polish-based project of a similar nature is PELCRA (Polish and English Language Corpora for Research and Applications) developed by the Department of English Language at Łódź University in cooperation with the Department of Linguistics and English Language at Lancaster University (cf. Lewandowska-Tomaszczyk 2001). One of the corpora built as part of the project is a learner corpus of ca. 3m words of spoken and written English as produced by Polish learners, which has served as a basis for a number of studies on learner English, including error analysis (e.g. Lewandowska–Tomaszczyk, Leńko-Szymańska and McEnery 2000).

The most extensive use of learner corpora in ELT has been made in producing learner dictionaries. The list of learner-corpus-based publications includes the *Longman Language Activator* (1993) the *Longman Dictionary of Contemporary English* (1995), the *Longman Essential Activator* (1997) and the *Cambridge Advanced Learners' Dictionary* (2003) (cf. De Cock and Paquot 2010). One of the newest and the most advanced in this respect is the *Macmillan English Dictionary for Advanced Learners* (2007) (henceforth *MED2)*, which was developed in close cooperation with the CECL, and strongly depends on data from ICLE (cf. Rundell and Granger 2007). The dictionary has a special feature: 100 boxes called "Get it right" (GIR), in which learners' attention is drawn to those aspects of English which are a source of problems and errors to learners worldwide. These sections were based on findings from ICLE, with the selection of material following the process described in detail below:

> MED2 is aimed at all advanced learners regardless of their mother tongue backgrounds. It was therefore decided that, to make the shortlist of flagged words or phrases, the errors associated with these items had to be shown to be both frequent and widespread (i.e. attested in data from learners from at least five different mother tongue backgrounds) in the learner corpus. Possible candidates for the shortlist were identified on the basis of, on the

> one hand, careful scrutiny of the 680,000-word error-tagged component of ICLE available at the time and, on the other, systematic comparisons of word frequency lists from the advanced learner corpus and the native speaker corpus used. Once identified, the possible candidates were subjected to rigorous analysis in the whole 3.5[million]-word ICLE corpus. The errors in the GIR boxes cover a number of categories including 'countability', 'register', 'verb patterns' or 'spelling' (De Cock and Paquot 2010).

Other learner-corpus-based features of the dictionary include six Grammar Sections and twelve EAP Writing Sections. Quite appropriately for a dictionary, its Grammar Sections focus on word grammar (problems like complementation patterns, countability, or quantifiers), which corresponds with the unique character of corpus findings. As for the EAP Writing Sections, they present those rhetorical or organization functions characteristic of academic English which learner-corpus analysis proved to be particularly challenging for learners of English as a second/foreign language (e.g. expressing personal opinions, expressing possibility and certainty, introducing topics and related ideas, or listing items).

Apart from dictionaries, reference grammars for learners have begun to depend on learner corpora to a considerable extent. This is the case with Carter and McCarthy (2006), where it is explicitly stated that a large learner corpus (albeit not identified by reference or name) was consulted in the process of writing the grammar, with the result that some elements in it are marked with a special symbol as "common areas of potential error" (Carter and McCarthy 2006: 16).

1.3.9. New prospects for analyzing interlanguage

Error analysis was widely criticized after its initial success in the 1970s. The charges against it are best summarized by Celce-Murcia and Schachter (1977: 446-447), in the list of the six weaknesses of EA presented below. Each of them is discussed here in view of the new opportunities that corpus-based methodology offers. It appears that most of the criticism leveled at EA does not apply to computer-assisted interlanguage analysis.

1.3.9.1. The analysis of errors in isolation

Traditional EA extracted the error from its context, and – what is even more objectionable – ignored instances of the correct use of a given form in the data analyzed. This is believed to be the main and heaviest charge against EA. With the corpus-based approach, errors retain their immediate context (if analyzed and presented in the KWIC format), allowing the researcher and then the reader to understand how the error arose. More importantly, corpus analysis permits the researcher to compare the instances of erroneous use with those where the form was used correctly. Indeed, only then is interlanguage analysis thorough enough to allow any generalizations.

1.3.9.2. The proper classification of identified errors

Classifying errors is the biggest difficulty for the EA researcher: on the one hand, there is the unique character of the study and of the collected data, which need to be accounted for in the most appropriate way; and on the other – the body of previous research where certain classifications have already been developed and used. Learner corpus work developed at the CECL is gradually becoming dominant in the field, which may lead to the universal acceptance of their nomenclature. A similar process took place with CLAWS (the Constituent Likelihood Automatic Word-tagging System), a tagset developed by the University of Lancaster and used for parts-of-speech (POS) annotation of the British National Corpus (Leech et al. 1994). CLAWS has now become the leading standard in the POS tagging of corpora. So far the CECL's error annotation system has not been fully standardized, and changes slightly from one research project to another (compare, for example Dagneaux et al. 1998 and Granger 2003), but the main principles have already been established: errors are classified in descriptive terms (linguistic taxonomy – see section 1.6.1.1.), with several main categories divided into subcategories depending on the formal classification of the language item affected by the error. There is one aspect of error classification which seems to be a source of a high level of confusion: the distinction between grammatical and lexical errors. These categories keep overlapping and always need to be specially defined and artificially demarcated. It would be highly beneficial for the field to follow the recommendations from

Halliday (1985), Sinclair (1991) and Lewis (1993) and abolish this division, introducing the broad category of *lexicogrammar* in place of the two or even three 'fuzzy' ones (with lexicogrammar sometimes introduced as an in-between category, e.g. Dagneaux et al. 1998). Introducing this third category, however, did not seem to solve the problem, as the ambiguity still remained: Dagneaux et al. (1998) report having to instruct their staff on classifying the use of the indefinite article in front of uncountable nouns (**an advice*) as either a grammatical error of article selection, or a lexical error of countable/uncountable distinction (the reader is not actually informed which decision was recommended). Merging grammar and lexis into lexicogrammar would help resolve such debates, allowing for the inherent ambiguity of the language system.

1.3.9.3. Statements of error frequency

EA was accused of offering only vague assessments of the frequency of occurrence for errors in less formal studies; more sophisticated ones, which took into account *relative* error frequency – calculated against the total of obligatory contexts in the sample – were criticized here for being unable to include *optional* contexts. A good example of such a study is the one by Faerch (1978, as cited in James 1998: 18), who proposed replacing EA with Performance Analysis, which took into account both errors and non-errors. Although it indeed was an improvement on EA, it still depended on the learner actually having used a given form. The problem with such an approach is that it fails to account for the common strategy of *avoidance*: when learners find a form difficult (either because of its internal complexity or because of the complex rules of its use), they refrain from using it, and instead seek 'safer', less error-prone forms to express themselves. The corpus approach can minimize charges against such studies by providing statistics on avoidance, obtained by comparing IL frequency statistics with NS corpus statistics for a particular language form. This should partially satisfy the need for including optional contexts in the research though – admittedly – they do so indirectly.

1.3.9.4. The identification of points of difficulty

The fact that EA focused on errors ("what the learner will do"), assuming that they are the only indication of learner difficulty, and overlooked the phenomenon of *avoidance* ("what the learner won't do and why") was a serious drawback. In learner-corpus analysis, this is no longer an issue: as mentioned above, comparative analysis of native speaker corpora and learner corpora provides useful data on avoidance at various levels of language use.

1.3.9.5. The ascription of causes to systematic errors

The main difficulty within EA was establishing whether a particular error was *interlingual* or *intralingual*. Most of such judgments were based on researchers' intuitions or findings from contrastive analysis, which many scholars found insufficient. Corpus methodology opens new opportunities in this area, in the form of contrastive interlanguage analysis (CIA) (Granger 1998a), by means of which various learner corpora of different L1 backgrounds are investigated to find whether a given non-native language form is characteristic of learners who share a particular L1, or whether it constitutes a more widespread phenomenon and should be considered an intralingual problem. It must be noted that such analysis requires very strict criteria of corpus construction to be followed, so that no external variables (differences in proficiency level, topic, genre, style, age, etc.) would obscure the results. There are, however, corpora designed for this very purpose, such as ICLE or the Longman Learner Corpus (LLC). Granger (1998a) emphasizes, however, that CIA should always be supported by reliable contrastive analysis data to exclude the possibility that the various L1s under investigation do share features which may affect the learners' L2 performance in a similar way. The same point was raised in the original criticism of EA by Celce-Murcia and Schachter (1977).

1.3.9.6. The biased nature of sampling procedures

The data collection methodology of traditional EA resulted in samples very limited in their amount and scope, mainly because of its labor-intensive character. This is not an issue with learner-corpus studies. The common size of a learner corpus used in research is 200-300 thousand words, which is sufficient unless the study focuses on forms of particularly low frequency. Corpora are usually built in such a way as to provide data relevant to the research question in terms of a variety of background languages (if the study is comparative) or other characteristics of the subjects. Issues of quantity cease to pose a challenge; what requires care and planning is the design of corpora. For generalizations to be made about language acquisition and its various aspects, the corpora must be large, while other features (such as multiple L1s, various age groups, proficiency levels, and learning situations) must be strictly controlled; the information about these features must be made available to the researcher so that s/he can control them in corpus queries. There is another 'charge' included in this category: that EA depended only on *performance* data, which may be biased in some way, and which was wrongly identified as representative of learners' interlanguage *competence*: "There is the ever-present danger of treating performance data as if they were the only and ultimate truth" (Celce-Murcia and Schachter 1977: 449). Corpus studies by their very nature depend exclusively on performance data and assume this to be the only accessible source of information about learner language. The large size of corpora eliminate the risk, which indeed existed with EA, that an important feature may not appear in the data by mere coincidence, or even purposeful resistance, of the subject(s). With corpora, the language sample is usually collected under no pressure from the researcher and with no observable agenda on their part, so such bias is highly unlikely. With EA this may not have been the case.

The discussion above should make it clear that there is a significant qualitative difference between traditional error analysis and computer-assisted interlanguage analysis. The latter gives a deeper and more accurate insight into features of interlanguage, with much higher statistical accuracy and a broader scope of inquiry. Certain well specified techniques of corpus analysis allow researchers to obtain hard quantitative data on various types of errors and other aspects of interlanguage, less accessible by means of traditional error analysis. Learner-corpus analysis does have its own challenges and

weaknesses, but these can be overcome with enough resources, or at least taken into consideration when generalizations and conclusions are made.

Errors remain an important source of information about the process of language acquisition and about interlanguage as such. They are also teachers' and learners' daily reality, and often a source of frustration for both. This is why researchers into language acquisition are still intrigued by how and why language errors arise and want to explore these questions, difficult as they may seem. Mainstream language acquisition research had moved away from error analysis for some time, until in the 1990s it was revived by researchers who saw new opportunities offered by multilingual corpora. Error analysis may still bring valuable results, learner errors being too important a factor in language learning to be ignored. Now, however, it is approached from a different angle than in the 1970s and is informed by extensive learner corpus research that has been undertaken since that time. Even the newest publications acknowledge that error analysis is not a thing of the past: "error analysis continues to have a role to play in remedial approaches to teaching of writing" where researchers "examine L2 learner errors as a basis for deciding what L2 features to teach" (Ellis 2008: 65). This is exactly the role that error analysis (or what is now more adequately called interlanguage analysis) plays in this study.

1.4. Transfer

Awareness of cross-linguistic influence is as old as people's recognition of foreign accents. It emerged in language acquisition studies in the 1940s and 1950s, with such names as Charles Fries and Robert Lado. Their theories of transfer were based on behavioral psychology, dominant at that time, especially in the United States. Another important contributor to the current understanding of the concept is Weinreich ([1953] 1968) with his idea of "interlingual identifications" in studies of bilingualism. The very term *transfer* originates from psychology and refers to "the persistence of, or resort to, already existing behaviour in a functionally new behavioural activity" (Corder [1978] 1981: 99). This general concept, when applied to language acquisition, caused much controversy over how it should be understood, and whether it should be used at all. One of its most recent and satisfactory definitions was given by Ellis (2008: 351): "Lan-

guage transfer refers to any instance of learner data where statistically significant correlation (or probability-based relation) is shown to exist between some feature of the target language and any other language that has been previously acquired". Owing to its reference to statistics, this definition surmounts many methodological problems which other attempts, discussed below, failed to overcome. Ellis, like Odlin (1989), uses the word *transfer* interchangeably with *cross-linguistic influence* (CLI), the term proposed by Sharwood Smith and Kellerman (1986) as "theory-neutral" and therefore more universal. The same practice has been adopted in this study.

1.4.1. Misconceptions

The concept of transfer being one of the most contentious issues in SLA research, its presentation here begins with eliminating some of its most common misinterpretations. A very clear overview of controversies related to transfer was presented by Odlin (1989). Four major reservations were made about what is often assumed about transfer and what clearly is not part of the current understanding of the term within the field of language acquisition research. First, Odlin makes it clear that transfer must not be identified with behaviorism and habit formation. The fact that the concept was introduced by linguists associated with behaviorism, whose standing in linguistics was severely undermined in the 1970s, contributed largely to its lack of appeal at the time. Transfer as such, however, is *not* exclusively a behaviorist idea and is part of many other theories of learning formulated before and after behaviorism.

Another misconception that Odlin (1989) dismisses is that transfer is identical with interference. This misunderstanding is probably rooted in the fact that linguists seem more interested in those aspects of transfer which lead to errors and increase the difficulty of learning a second or foreign language. These are often covered by the term *negative transfer*, or *interference*. Odlin recommends using the former, because it emphasizes the opposite phenomenon, *positive transfer*, by means of which knowledge of one language facilitates the learning of another. This is believed to be the case especially if the languages in question are genetically related.

The next claim that Odlin (1989: 26) rejects is that "transfer is simply a falling back on the native language". This is an idea coming from Krashen (1983), who reduces

transfer to 'padding', i.e. a *communication strategy* adopted by the learner to resort to an L1 rule when s/he does not have sufficient means of expression in the target language. Krashen believes such a strategy does not facilitate the learner's progress (is not a *learning* strategy) but merely helps the learner survive. This approach has become known as the *minimalist position* on transfer. What such a view ignores, according to Odlin, is that speakers of languages similar to the target language do have an advantage in comparison to those whose languages bear little resemblance to it. The gain does not necessarily have to be a straightforward rule known from L1, but general similarity in the lexicon, writing systems, features of discourse, or other aspects of language, and interaction between these different factors. These similarities are beneficial not only in language production, but also in comprehension. Therefore, limiting transfer to an ad-hoc production strategy does not account for its full contribution to language learning.

Finally, discussions of transfer often assume that the influence comes from the learner's native language. This is a major overgeneralization, as any previously acquired language(s) can have an effect. What is more, the direction may be reverse: apart from the more commonly recognized transfer from a native or previously learned language to a target language, formally referred to as *substratum transfer*, an effect of L2 on L1 is also possible, where *borrowing transfer* takes place. The latter phenomenon was recognized and distinguished from the former by Thomason and Kaufman (1988) – in their definitive work on language contact – and in extreme cases it may lead to language attrition. This is often observed in migrants, who incorporate elements of their host country's language, especially lexis, into their L1. An illustration of this phenomenon is this sentence from a Polish immigrant in America in a comment made on a fashion website:

> *widze ze ludzie zaczynaja byc snobistyczni i wydadza pieniadze nawet za worek* ***garbeciowy*** *aby bylo napisane LV*[8] (original orthography preserved)
> 'I can see that people begin to be snobbish and will even spend money on a garbage bag as long as there is an LV label on it.'

Worek garbeciowy is a case of a borrowing transfer of *garbage bag* into Polish, and would not be understood by most native speakers of the language. Borrowing transfer

[8] http://www.figa.pl/Markowa-torba-na-smieci-za-2-tysiace-dolarow-a65/2

can operate at the level of syntax too, as in this sentence from a comment on another blog:

> ***Ja*** *znam sporo osób polskiego pochodzenia* ***tutaj***.[9]
> 'I know many people of Polish descent here.'

The sentence does not strike the Polish reader as gravely incorrect, but clearly does not sound natural. It has an overt first person singular subject, which would normally be ellipted, and the adverbial of place in the sentence-final position, which is not 'default' in Polish. (It would be more natural to place it after the verb in this case.) Both of these are typical features of English syntax.

1.4.2. Definition of transfer

After specifying what transfer is not, Odlin goes on to provide his working definition of it, which has been very commonly quoted in discussions on transfer and is accepted as binding here as well: "Transfer is the influence resulting from similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired" (1989: 27). This definition does not eliminate the problems of experimental research, as it does not offer much guidance in relation to how the occurrence of transfer might be clearly recognized. Odlin does, however, make suggestions as to what steps should be taken in order to make such research more reliable. Ideally, instances of transfer should be diagnosed as such in two independent ways: by means of a contrastive analysis of L1 and L2 and by comparing the performance of speakers of different L1s as learners of the same L2. As for the former, Selinker makes the following statement: "CA is the best place to begin language transfer studies since structural congruence (or at least partial structural similarity) is most probably necessary, though not sufficient, for many of the claims regarding CLI" (Selinker 1992: 208-209). In the latter type of diagnosis, differences in various L1 speakers' use (and misuse) of L2 would indicate that particular problems result from the different language

[9] http://www.resvaria.net/2009/03/04/polacy-w-chicago-prawda-czy-mity

backgrounds of the groups. In further discussion, Odlin (1989) stipulates that such comparative studies are not error proof either: it may happen that different L1s share particular features which, as a result of transfer, generate similar errors in L2. It may be equally likely that the L2 feature in question is so unique that speakers of other languages find it equally challenging and produce similar numbers of errors when attempting it. A good example of such a feature in pronunciation could be the articulation of interdental fricatives /ð/ and /θ/ in English. The sound is so rare among phonological systems that many learners will find it equally challenging.

1.4.3. Structural differences and the hierarchy of learning difficulty

Systematic differences between L1 and L2 can affect the process of second language acquisition in a variety of ways. For example, the hierarchy of learning difficulty in Table 1 below places an absence of an L2 feature in the learner's L1 as the highest indicator of difficulty.[10] The hierarchy was designed by Stockwell et al. (1965: 284) as a very complex system, and then reconstructed in Ellis (2008: 360), where distinctions between optional and obligatory elements included in the original were omitted. This produced a much more straightforward and transparent system presented below.

As Table 1 demonstrates, the common belief that a large difference between two corresponding elements in two languages will automatically indicate higher difficulty for the learner is an oversimplification. Similarity can cause problems as well: second on the list is the situation where one element in L1 has two or more corresponding elements in L2, which results in very common errors of underdifferentiation. There are correspondences between the two systems, but learners fail to recognize distinctions which are not made overtly in their L1, and so use one form only, or both, but without making appropriate distinctions between them. An example most familiar to Polish teachers of English will be their learners' use of the Past Simple tense instead of the Present Perfect, for these very reasons: Polish uses its past tense for most contexts in which English speakers would apply either of the two tenses.

[10] The label "Feature absent from L1" appears frequently in the error analysis performed for this dissertation, especially with article errors.

Table 1. Hierarchy of learning difficulty (adapted from Ellis 2008: 360) [11]

Type of difficulty for L2 category	L1: English	L2: Polish	L1:Polish, L2: English Examples
New	Ø	x	articles, tense sequence, auxiliary inversion
Split	x	x/y	'na' is 'on', 'at' or 'for', etc.; 'który' may be 'that', 'which' or 'who', 'whose', etc.; 'swój' is 'my' or 'mine', 'your' or 'yours', 'her' or 'hers', etc.
Absent	x	Ø	case inflection in lexical nouns and adjectives, grammatical gender, Pronominal Subject Deletion
Coalesced	x/y	x	Polish T/V forms of address vs. 'you', lexicalized gender distinctions ('nauczycielka'/'nauczyciel' vs. 'teacher')
Correspondence	x	x	simple and complex comparative forms

The classification may be compared with what the present study observed to be very common and persistent errors in Polish learners of English. For example, the high frequency of indefinite article omission seems to confirm the status of the top-level category; the indefinite article is absent from the Polish language system, whereas it is obligatory in front of countable nouns in English (unless other determiners occur). This difference between the two languages can be held responsible for numerous errors in fairly advanced learners' output.

On the other hand, an error which is also quite frequent, namely the use of a comma in front of *that* as a relative pronoun or conjunction, would be classified as the fifth rank in the above hierarchy: obligatory in Polish and absent from English (unless *that* is preceded by a parenthetical expression, which needs to be set off with commas). An explanation of this particular problem's persistence may lie in a special case of a transfer of L1 training: teachers of Polish instill this rule of punctuation very deeply in their students, which may result in the same use of the comma in English.

There are two important concepts connected with transfer and how particular attributes of L1 and L2 can affect a foreign language learner's success. These are *markedness* and *prototypicality*. Markedness relates to features of L1 and L2 which are in some ways 'special' or 'unique'. According to markedness differential hypothesis

[11] The table has been adapted in two ways: first, Polish rather than the original Spanish examples are provided, and second, the sequence has been changed to make it match the original source of the hierarchy (Stockwell et al. 1965: 284).

(Eckmann 1977) such features are believed to be less easily transferred from L1, especially if the corresponding feature in L2 is unmarked (i.e. there is no form in L2 dedicated to serve a corresponding function); and the reverse – unmarked L1 features are easily transferred, especially if the relevant L2 feature is marked (cf. Ellis 2008: 385). An example of such a situation, examined by Eckmann (1977), is that of the asymmetrical pattern of voicing in word-final stops in German and English. He argues that the English feature of (partial) voicing in this position is marked (less 'normal') and the German lack of voicing – unmarked. That is why German learners have problems producing voiced word-final stops in English (an unmarked feature of devoicing is transferred in place of a marked feature in the target language), but English learners of German have no problems articulating voiceless word-final plosives (L1 marked feature not transferred to L2 to replace an unmarked feature). These observations are not, however, by any means universally accepted by linguists.

To explain the concept of *prototypicality,* Ellis (2008: 389) discusses research by Kellerman (1983), "who sought to demonstrate that learners have perceptions of the structure of their own language, treating some structures as potentially non-transferable and others as potentially transferable, and that these perceptions influence what they actually transfer". In other words, some forms in L1 are believed to be 'core' (more prototypical, semantically transparent and universal) and others 'non-core'. The latter are believed to be less willingly transferred, presumably because learners perceive them to be less likely to be similar in another language. Kellerman's research focused on lexical forms and semantics, but his later work is reported to have made similar findings in conditional clauses, i.e. syntactic forms (Kellerman 1989). As Jarvis and Pavlenko (2008: 188) observe, *prototypicality* and *markedness* are both connected with speakers' intuitions about their L1 forms, how prevalent (i.e. prototypical and unmarked) they are and, consequently, whether they can be transferred to another language. They also account for the fact that language influence is often asymmetrical (one language may transfer particular features to another, but transfer does not occur in the reverse direction). Jarvis and Pavlenko make it clear that these constraints on transfer are far from absolute.

1.4.4. Non-structural constraints on transfer

Differences in the degree of L1 influence are recognized in connection with the numerous non-structural factors dependent on individuals or the context in which they learn the language. Some, like for example personality, motivation or language aptitude are involved in any aspect of foreign language learning and are largely self explanatory. Others, like *age* or *L2 proficiency*, have caused some controversy among researchers and deserve more attention.

1.4.4.1. Age

From the SLA perspective, it is age that is most commonly (though not universally) perceived as an inhibitory factor, and so most researchers believe that the earlier children begin to learn a second or foreign language, the better the results will be. This is especially the case with pronunciation: younger children have a higher rate of success in achieving native-like pronunciation (Seliger et al. 1975). As for other aspects of language, Corder ([1976] 1981: 74) observes that "in the case of older children, either in a formal or informal learning setting, the influence of the mother tongue or other known languages becomes more evident, until, when we are dealing with adults, particularly if they are educated, interference seems to be strongest". The explanation of this phenomenon may be that since very young learners still have access to UG (see below), they can depend on it *and* the L2 input to build their IL. This mirrors the processes of L1 acquisition. Older learners, on the other hand, view new experiences in the perspective of the knowledge they have acquired previously, and so they attempt to construct the target language system through categories developed earlier for their mother tongue. L2 is perceived through the categories developed for the purposes of L1. Wherever there is a gap in L2 knowledge, the learner will resort to forms or rules that are available in L1. There are, however, other studies which question these claims, suggesting that there may be other factors responsible for research results like those in Seliger et al. (1975), such as length of residence in the L2-speaking country (e.g. Purcell and Suter 1980). Also, mature learners benefit from their advanced cognitive and literacy skills, which may enhance their L2 performance. Generally, the effect of the age factor in second

language acquisition is unresolved; there is a possibility that some elements of linguistic competence are acquired more effectively at a younger age, while others – later.

1.4.4.2. Target language proficiency and task type

Opinions concerning the relationship between proficiency and L1 interference vary, too. Taylor (1975) argued that learners rely on transfer more extensively at the lower levels of L2 proficiency, because it is then that the gaps in their knowledge of the L2 system are the broadest. Hence the amount of interference is biggest at that stage. This may be connected with what Corder stipulates in one of his papers: "The more demanding the communicative activity required of a learner (...) the greater the amount of borrowing in general and hence unsuccessful borrowing (interference error)" (Corder [1978] 1981: 100). A more general conclusion can be made, then: the type of task the learner performs may affect the level of interference. Taylor's (1975) claims were based on a translation task, in which more advanced students could rely on their knowledge of the target language system and had more grounds for drawing some analogies when looking for solutions. The less advanced learners did not have such a strong foundation to depend on, and had to resort to their L1 for help. As a result the more advanced students' errors were recognized as mostly intralingual, whereas the less advanced students' errors – interlingual. The effect, however, may have been enhanced by the level of difficulty of the task. Odlin (1989) makes an important reservation with reference to Taylor's findings: some types of negative transfer cannot operate until learners have developed a command of more advanced features of L2, and only when that happens, can L1 influence become visible. For example, errors identified in this study, connected with the use of gerunds with *by+NP* (**teaching by parents*) would not appear at the elementary or lower-intermediate level, because students would not be likely to build such complex structures. Also Kellerman (1983) questioned the common belief that transfer errors are more common in beginners. Odlin concludes that "transfer can occur whenever a structure seems new, and therefore problematic, to a learner. While Taylor's study is an important indicator that proficiency may interact with transfer, the relation between target language ability and native language influence is likely to be a complex one" (Odlin 1989: 134).

Translation is often believed to generate an enhanced quantity of interference errors: Duley et al. (1982: 110) claimed that "the proportion of interlingual errors changes with the elicitation task, translation in particular. Translation tasks artificially increase the L2 learner's reliance on first language structures, masking processes the learner otherwise uses for natural communication". It is true that translation tasks generate more errors than, for example, writing, but as Meriö (1978) observed, this results from the very nature of the task. When writing, students use only those structures and forms of which they feel relatively confident. Stipulations that translation elicits more *interference* errors than other tasks have been challenged in some experimental studies. Mattar (1999), for example, analyzed Arabic learners' interference errors in the use of English articles. The data for that study were elicited in two ways: through translation and with a gap-filling task; it was actually the latter technique that proved more error-prone. On the whole, there are many experimental studies that prove the superiority of CA and translation techniques over other types of form-focused instruction and their higher effectiveness, e.g. Sheen (1996) or Laufer and Girsai (2008). Strong support for bringing CA and translation back to the language classroom was also expressed by James (2005). A very recent and very well argued voice on the issue comes from Cook (2010), who sees translation as a way of developing intercultural competence and language awareness. He also maintains that translation gives students more confidence and supports their autonomy.

1.4.4.3. Literacy

According to Odlin (1989), native language literacy (understood as an ability to read and write) can have a positive effect in second language acquisition. This may not be, however, straightforward positive transfer but rather a matter of *transfer of training*: the problem-solving skills acquired in the process of L1 literacy development during a learner's formal education can enhance L2 learning and make it more successful. This seems especially true of the lexical aspect of language: it is impossible to develop a rich L2 vocabulary unless one's L1 stock of words is broad enough. On the other hand, L1 literacy can be an inhibiting factor for developing writing and reading skills in L2 if the two languages' writing systems are very different.

1.4.4.4. Social background

The effects of transfer can be strongly affected by the social context in which the learner lives. For example, in post-colonial countries like India or Singapore, speakers of English regard their version of it as a fully developed language variety in its own right. They simply may not aspire to speak standard British or American English, because by contributing features of their L1, they may be marking their national or ethnic identity and be proud of it. A similar situation could be observed in Ireland, where "the differences between Hiberno-English and British English became a source of nationalist pride for some leading literary figures. (…) What is usually considered negative transfer can thus persist sometimes and achieve a favorable social status" (Odlin 1989: 149). By analogy, immigrants to English-speaking countries need to make a choice: they can either emphasize their ethnicity and intentionally preserve L1 features in their English, or they may want to become wholly assimilated into the host culture and language, and avoid transfer as much as they can. These choices depend on the strength of family traditions, attitudes within their ethnic community, level of education, length of stay in the host country, political views, and many other individual characteristics.

There is also a matter of the amount and quality of input learners receive. If they do not have easy access to native speakers of standard English but are immersed in the community that speaks its ethnic variety (a common situation in *second* language acquisition), or can only use English with their peer learners (typical of *foreign* language learning), then transfer is more likely to have a strong effect. The latter phenomenon was an object of a study on a bilingual immersion program in Canada (Selinker et al. 1975), which "suggests that the social context, a relaxed setting of learners speaking with each other, may well have encouraged cross-linguistic influence" (Odlin 1989: 104). The problem is more broadly addressed in section 1.5. on transfer errors and fossilization.

1.4.4.5. Linguistic focusing

A factor directly connected with ethnic identity is linguistic focusing. According to Odlin (1989: 144), "focusing involves: (1) some awareness of belonging to a group; (2)

considerable awareness of linguistic and other norms that distinguish one group from another; and (3) adherence to and enforcement of such norms". The phenomenon of the powerful influence of ethnic identity mentioned above is, then, a typical example of linguistic focusing from the point of view of the 'outsider' community. There is, however, the other side of this problem: the host community may feel the need to protect its standard language variety from foreign influences and discourage its members, as well as newcomers, from introducing foreign elements into their language. Such tendencies are especially strong in France, where the *Académie française* ('The French Academy') guards the purity of the French language against foreign, especially English, influences.[12] Learners of languages whose native speakers share these sentiments are under considerable pressure to control negative transfer in order to prevent unfavorable reactions from their target language community. Fortunately, these attitudes are not usually directed against foreigners, but against other native speakers who simply prefer to choose foreign words instead of their original, native equivalents. This is not a common problem among native speakers of English, who are accustomed to foreigners using their language without much inhibition. One problem that does sometimes occur is preventing American English from permeating into British English. This, however, is of less concern to learners of English as a foreign language.

1.4.4.6. Learning context: formal instruction vs. natural setting

James (1994) observed that students learning a language in a formal setting (the classroom) were more likely to resort to transfer as a strategy than those who acquired a foreign/second language in the natural setting. This can be easily explained by the fact that classroom instruction usually offers insufficient L2 native speaker input for the learner to acquire the target language system, in which case the learner resorts to the system(s) s/he already knows, usually L1, as the support. Such situations may lead to results which confirm Hammerly's (1991: 21) *mock* motto "Practice makes permanent". As he explains, "Practice of the wrong kind or at the wrong time, and practice without effec-

[12] See http://www.academie-francaise.fr/dire-ne-pas-dire/neologismes-anglicismes

tive feedback makes imperfection permanent. Communicating in a second language with many errors makes the faulty rules underlying the errors permanent".

1.4.5. Transfer: constraint, strategy or process

To summarize the above discussion of transfer, it may be useful to see the development of the concept in three different directions, which to some extent may be seen as developing chronologically. The first stage is that of transfer as a *constraint on learning*. Within the behaviorist tradition the constraint lies in the habits carried over from L1 to L2, which prevent the learner from successful learning of new language behavior. The constraint view of transfer is not, however, unique to behaviorism. The Hypothesis Theory, which belongs in the mentalist framework (contradicting behaviorism), also adopts the view that transfer is a constraint (Schachter 1992). Here, however, it is not a constraint on habit formation in language learning but on the learner's hypothesis testing involved in formulating rules: "[transfer is] the set of constraints that one's previous knowledge imposes on the domains from which to select hypotheses about the new data one is attending to" (Schachter [1983] 1992: 39). The previous knowledge referred to here includes knowledge of one's mother tongue. The domains mentioned in the quotation are natural groupings of all the hypotheses that the learner has at his/her disposal (depending on previous experience), and which share certain characteristics and therefore narrow down the range of choices. According to Schachter, the hypothesis testing may be either conscious or subconscious.

Alternatively, transfer can be seen as a strategy (e.g. Krashen 1983 and Corder 1992). Since Krashen's views on transfer have already been discussed, suffice it to say that he sees it as a communication strategy that learners resort to whenever their IL proves insufficient for their communicative needs ('padding'). Corder subscribes to a similar view, referring to the phenomenon of *borrowing*, which he understands to be a performance phenomenon, "the use of items from a second language, typically the mother tongue, particularly syntactic and lexical, to make good the deficiencies of the interlanguage" (Corder [1983] 1992: 26).

The third major option is the understanding of transfer as a matter of process. It is attributed mainly to the cognitive view of language and is supported by such linguists

as Kohn (1986) and Odlin (1992), who employ Weinreich's (1953) concept of interlingual identification. Learners are believed to recognize similarities between those language systems they know and those they are becoming familiar with. These are employed both in L2 learning and L2 production:

> Transfer as a process is part of the learner's interlanguage behaviour, which includes not only the creative transformation of input data into interlanguage knowledge, but also the use to which the learner puts this knowledge in an attempt to produce meaningful and/or correct output. (...) According to the proposed distinction between knowledge and the use of knowledge in output, transfer can assume one of the two shapes: it can be a learning process or a production process (Kohn 1986: 22).

Along the same lines Odlin (1992) conceives of transfer as a selection process in which learners see *some* L1 forms and processes as more transferable than others, depending on such factors as frequency of occurrence and language distance *as perceived by the learner*. These are psycholinguistic criteria relating to Kellerman's concept of prototypicality discussed above.

Seeing transfer as a process has an interesting theoretical consequence, noted by Gass and Selinker (1983: 6), who observe that there is one process of transfer, which, when observed in the learners' language production, renders IL forms which can be recognized as 'positive', 'negative' or 'neutral'. In other words, one process may bring positive results (if L1 influence is successful and the IL form is consistent with L2 norms), negative results (if L1 influence generates errors), or neutral results – if no pattern emerges and the resulting IL forms are either errors or non-errors.

A compromise between the three competing views presented above was offered in Selinker's conclusion to his considerations on transfer:

> Language transfer is best thought of as a *cover term* for a whole class of behaviours, processes and constraints, each of which has to do with CLI, i.e. the influence and use of prior linguistic knowledge, usually but not exclusively NL knowledge. This knowledge intersects with input from the TL and with universal properties of various sorts in a selective way to help build IL (Selinker 1992: 208).

1.4.6. Transfer and Universal Grammar

More recent studies connect transfer with the concept of Universal Grammar. Nativists assume that language is dependent on an inborn system, and some of them posit the

existence of Universal Grammar, a set of universal principles that are common to all languages and constitute the basis of the native speaker's competence in language. Apart from these, current theories within generative grammar assume that UG contains parameters, which have several possible values. The values characteristic of particular languages are called settings, and are selected in the process of acquisition on the basis of language input. Principles and parameter settings are at the core of every language and its grammar.

An aspect of UG that is the most relevant to language acquisition is that Universal Grammar "is taken to be a characterization of the child's prelinguistic state" (Chomsky, 1981: 7). In other words, it is an innate language faculty, which may be involved not only in acquiring a first language, but also a second language and more. Different variants of relations between L1 and UG are proposed, some of which take into account L1 influence on the acquisition of other languages later in life, while others exclude such a possibility (Gass and Selinker 2008).

The two opposing views concerning interaction between L1 and UG are the Fundamental Difference Hypothesis on the one hand and Access to UG Hypothesis on the other. The former claims that the only access adults can have to UG is indirect, through their mother tongue. Connected with this is the assumption that L1 acquisition is fundamentally different from learning other languages later in life, because the former involves development of the very concept of language and (subconscious) knowledge of its main principles, which the process of second language acquisition then employs and from which it benefits. The process is also supported by general problem-solving skills that the individual has managed to develop before undertaking the task of learning another language. All this makes L1 and L2 acquisition two very dissimilar processes.

The position referred to as Access to UG Hypothesis contradicts the view outlined above. It actually represents a variety of options, defined by White (2003), which are all based on the assumption that a second language learner has at least some access to UG. They differ in the degrees of influence that L1 and UG are supposed to have and are briefly characterized as follows (White 2003: 61 ff.):

L1 as the base

- *full access/full transfer*: The point of departure is L1, but the learner has also full access to UG if L1 proves insufficient.

- *Minimal Trees Hypothesis*: L1 and UG are both available, but L2 functional categories need to be formed by the learner on the basis of L2 input as the L1 grammar and UG available for transfer at the initial stages of learning are very rudimentary and have no functional elements.
- *Valueless features*: L1 is the primary point of departure, but generally transfer is weak. Both lexical and functional elements are available, but they are too weak to be transferred.

UG as the base

- *The Initial Hypothesis of Syntax*: The starting point for second language acquisition is UG.
- *Full Access/No Transfer*: L1 and L2 acquisition are totally disconnected, and proceed in a similar fashion, which means the L2 learner can be as successful as the L1 learner.

Considering the relatively low success rate in foreign language learning, the last option does not seem very feasible.

White (1992) discussed in more detail how UG research sheds new light on some issues related to transfer and defined four areas which connect the two. These were:

- *Levels of Representation*: Generative grammar is known to distinguish between various levels of language structure, the two most commonly recognized ones being surface structure and deep structure. When analyzing transfer, researchers should consider its occurrence not only on the surface, but also in relation to the deep (underlying) structure. CAH concentrated on 'surface' differences between languages, but the CLI may be present at deeper levels as well.

> It is quite possible for languages to have superficially similar sentence types, which in fact stem from very different D- or S-structures. According to the traditional CAH, these superficial similarities would be predicted not to cause problems, whereas on a G[overnment] B[inding] account which proposes structural differences in their analysis, transfer effects might be expected. (...) Conversely, different surface forms can result from the same D-structure (White 1992: 221-222).

- *Clustering*: Parameter settings usually involve clusters of properties, which on the surface may not appear to be related. Learning a language involves setting and resetting parameters, and research indicates that when one parameter of the cluster is reset in interlanguage grammar, the others do not seem to be automatically reset. The cluster of features that has been studied most intensively is related to the omission of subject pronouns, a feature that some languages (including Italian, Spanish and Polish) allow (cf. White 1985, Liceras 1989).
- *Multi-valued and interacting parameters*: Many parameters in UG need to be reset when L2 is being acquired. Since it is very unlikely that they will be reset at the same time, the combinations of parameters in IL may actually be different from that of either L1 or L2.
- *Learnability*: This concept is connected with the notion of markedness discussed earlier. Some parameter settings are marked ('special'), others are unmarked. Those that are marked in L1 are believed to transfer less easily, and unmarked ones – more easily (Eckmann 1977). The UG perspective adds to these considerations the assumption that L1 acquisition takes place on the basis of positive evidence, but the processes of L2 acquisition are different in this respect and at least in some cases may require negative evidence, especially where rules are overgeneralized on the basis of L1 settings (Yip 1994). Thus negative evidence needs to be provided to foster the learnability of a parameter setting, especially when there is nothing in normal input that would indicate the IL form used by the learner is in some way inappropriate. Such negative evidence may be offered as explicit correction from the teacher or some remedial work performed in class; it seems that concordances and other DDL materials can also be used to this effect.

1.5. Interference errors and fossilization

The above discussion of transfer should make defining interference errors a relatively easy task. It could be generally said that interference (or transfer) errors are those errors which result from negative transfer. For the needs of the research project undertaken for this thesis, however, the definition must be more accurate, so as to provide a clear guidance and make classification of errors as unambiguous as possible. For this reason the

definition and understanding of transfer errors adopted here is going to be the one provided by Zobl (1980: 470), who argues that the genesis of developmental and transfer-related errors may be to some extent connected: "[B]oth types of errors have their origins in the processing of properties of L2 input, and (…) errors which seem to show influence from the L1 presumably begin as L2-dependent developmental errors which are subsequently reinforced by an L1 structure compatible with the developmental error". In this way the unavoidable ambiguity involved in distinguishing between developmental (intralingual) and transfer (interlingual) errors (Dulay and Burt 1974) can be solved – there are cases where the two processes do overlap, L1 being an additional source of confusion and complication in overcoming problems that are developmental in nature. Such a view of transfer errors explains problems of overuse, for example, which are otherwise difficult to account for.

There is a highly practical aspect to recognizing transfer errors: on the basis of research by Mougeon and Hébrard (1975), Scott and Tucker (1974) and by Agnello (1977), Zobl (1980) observes that L1-based errors have a stronger tendency to fossilize than developmental ones. The same conclusions were reached by Selinker and Lakshmanan (1992: 198), who proposed the multiple effects principle (MEP), according to which transfer and fossilization are strongly connected: "When two or more SLA factors work in tandem, there is a greater chance of stabilization of interlanguage forms leading to possible fossilization." They further suggest that language transfer is either a *privileged* (weak form of MEP) or even a *necessary* (strong form of MEP) condition for fossilization. According to the authors, a pedagogical consequence of the above principle is that fossilized structures resulting from MEP cannot be modified by means of consciousness-raising interventions.

Fossilization appears to be a potential end result of transfer errors, which is most relevant to L2 acquisition and the degree of its success; fossilization is also directly relevant to this study, which focuses on the errors that were found to be particularly persistent and require special remedial intervention. The fossilized status of these errors makes the task of eliminating them an even bigger challenge.

It seems that some aspects of language education in Poland may actually enhance the effects of transfer. Almost all classes are composed of native speakers of Polish, and with the emphasis on communication practice currently dominant in Polish schools and encouraged by the school-leaving examination system, learners of English,

for example, often practice Polglish (the Polish variety of English) rather than English. James says:

> It is (...) demonstrable that learners of any given FL who share the same L1 and have been taught under similar conditions with the same texts and syllabus do emerge speaking the same 'social' dialect of that FL. They encounter few problems understanding each other's utterances rendered in that dialect, but outsiders, including speakers of the target dialect, that is 'native speakers', find it unintelligible (James 1998: 16).

A similar point is made in a study on error evaluation by Davis (1983), who proved L2 native speakers to show less tolerance of errors resulting from L1 interference than non-native-speaker teachers, probably as a result of problems with comprehension. What is most surprising is that the native speakers were on the whole more lenient in their judgments of error gravity, so those errors must have been perceived as particularly grave. This seems to be evidence enough that interlingual errors should be addressed seriously both in research and in language pedagogy. After all, successful communication is the key criterion in evaluating a student's progress in their mastery of a foreign language. This is why teachers should focus on typical L1 features in learners' L2 production, prevent them from fossilizing and expose students to large amounts of native-speaker input. Most of all, learners must not be forced to use L2 before they are ready to do so.

As far as the proportions between interlingual and intralingual errors in L2 research are concerned, research results are notoriously inconsistent. Table 2 shows the range of percentages for the former – from a minimal level (Dulay and Burt 1973) to a straight half (Tran-Chi-Chau 1975 and Lott 1983):

Table 2. Percentage of interference errors reported by various studies of L2 English grammar (from Ellis 2008: 355)

Study	Percentage of interference errors	Type of learner
Grauberg 1971	36	First language German – adult, advanced
George 1972	33 (approx)	Mixed first languages – adult, graduate
Dulay and Burt 1973	3	First language Spanish, children, mixed level
Tran-Chi-Chau 1975	51	First language Chinese – adult, mixed level
Mukkatesh 1977	23	First language Arabic – adult
Flick 1980	31	First language Spanish – adult, mixed level
Lott 1983	50 (approx)	First language Italian – adult, university

It may be assumed that the most extreme results may have been affected by some unusual conditions of the experiments in which they were obtained or perhaps unique sets of criteria applied in their classification; if this is the case, it can be assumed that about one third of errors can be attributed to L1 interference. That proportion, however, does not mean that L1 is just one third of the challenge learners face in their acquisition of the target language. Because of issues connected with error gravity and fossilization discussed above, L1 is a factor of great importance, and neither teachers nor researchers should ignore it.

1.6. Error classification

Literature of error analysis is very rich and diverse. Errors have been studied from a variety of standpoints and for numerous purposes. Over the long period in which error analysis has been performed, a unified and standardized framework for analysis has never been devised, which is one of the frequent charges made against the field. The wide variety of classification schemes used in error analysis can be itself divided into categories, which may introduce some order into this confusing area.

1.6.1. Descriptive taxonomies

Ellis and Burkheizen (2005) identify and discuss two major classifications, both of them descriptive in nature: linguistic taxonomy and surface structure taxonomy.

1.6.1.1. Linguistic taxonomy

Linguistic taxonomies are based on categories derived from a descriptive grammar of L2: "basic sentence structure, the verb phrase, verb complementation, the noun phrase, prepositional phrases, adjuncts, coordinate and subordinate conjunctions and sentence connection. More delicate categories relating to each of these can then be developed" (Ellis and Barkhuizen 2005: 60). Error analysis studies have applied a variety of such

categories, and often devised multi-level systems, aiming to account for the multitude of possible ways in which learners failed to conform to L2 rules. Such an approach is very practical, as it may have a direct application to language pedagogy, usually organized around grammar and structure. The disadvantage, though, is that it is based on the TL rules and categories, and does not recognize the fact that interlanguage – as a language system in its own right – has its own unique grammar. This problem was recognized as a "comparative fallacy" by Bley-Vroman (1983). James (1998) argues, however, that L2 learners are naturally assessed in terms of L2 rules and norms, and use them as points of reference, so linguistic taxonomy seems to have psycholinguistic validity. Studies employing linguistic taxonomies include Dušková (1969) and Burt and Kiparsky (1972). Lennon (1991) lists different types of errors classified in terms of categories of the language items involved: morphological, prepositional choice, article errors, NP pre- and post-modification, clause linkage, word order within the clause and sentence structure, lexical choice, and tenses. He also includes a possibility of blends between two attempted structures: "it may be an oversimplification to assume that a single target necessarily underlies a particular error" (Lennon 1991: 190).

A special case in the context of this project is the error tagging system developed for the needs of computer aided error analysis within the framework of the International Corpus of Learner English (Dagneaux et al. 1998) at the University of Louvain. As in many other cases, the tagging system is hierarchical and involves the following seven major categories and coding letters for them: form (F), grammar (G), lexicogrammar (X)[13], lexis (L), word redundant/word missing/word order (syntax) (W), register (R) and style (S). Apart from these, there are some sub-categories for more narrow distinctions, e.g. GV, (grammar: verb), GN (grammar: noun), and still more narrow subcodes within these GVAUX, GVT, GVN (auxiliary, tense and agreement errors, respectively). Granger (2003) presents a similar system developed for FRIDA (French Interlanguage Database), except that the nine top-level categories (the two extra ones being PUNCTUATION and TYPO) are named *domains* (a slightly confusing nomenclature in view of Lennon's (1991) totally different concept of domain – see p. 83). The learner corpus used for the present study does not have any error annotation, as it was one of the assumptions that it should be of the simplest possible kind, which any teacher can build

[13] The category of "lexicogrammar" is used differently in this study; for details see p. 88.

and use in a similar fashion. Still, the learner corpora compiled by academic institutions can have the feature of error annotation, and such corpora have opened new possibilities not only in error analysis but also in other areas of applied linguistics. These corpora are not, however, available to average teachers for either technical or financial reasons; their main use is in research or in writing language-teaching materials and language tests.

1.6.1.2. Surface structure taxonomy

Surface strategy taxonomy introduced by Dulay et al. (1982) is a straightforward system which depends on "the ways surface structures are altered in erroneous utterances/sentences" (1982: 61). The categories of analysis include:

- omission;
- addition (with regularization, double marking, and simple additions as sub-classes);
- misformation (with regularization, archi-forms and alternating forms);[14]
- misordering.

James (1998: 106) adapted this classification, criticizing and eliminating some of its subcategories, and preferred to call it Target Modification Taxonomy, arguing that reference to surface structure would indicate that there is an alternative *deep structure taxonomy*, which is not true. The original name of the taxonomy has been more popular with SLA researchers, however. James classified errors into the following five groups "based on a comparison of the forms the learner used with the forms that a native speaker (or 'knower') would have used in the same situation" (James 1998: 111):

- omission;
- overinclusion;
- misselection (wrong choice not wrong form);
- misordering;
- blend, which accounts for situations where an IL form is a result of the learner combining two L2 forms into one.

[14] There is some confusion in literature concerning this category. The original authors used the term *misformation*, as above. However, some researchers (e.g. Ellis and Barkhuizen 2005, Ellis 2008) use *misinformation* instead, even when referencing Dulay et al. (1982).

This set of categories was used as a secondary set of labels in the analysis of data collected for the present study. The two systems are independent of each other.

James raises an interesting question as a follow-up to the presentation of his Target Modification Taxonomy. He quotes research by Stemberger (1982) who analyzed native speakers' syntactic errors in speech and found the following order of relative frequency among the corresponding categories:

MOST……. Misselections > Blends > Deletions > Additions …….. LEAST
frequent frequent

Figure 5. Relative frequency of native speakers' speech errors (after James 1998: 114)

In connection with this, James wonders whether the sequence would be similar in learners' errors. Transfer errors in the essay database built for this study were assigned to the five classes, and the following order emerged:

- misselection – 176
- omission – 85
- misordering – 16
- overinclusion – 13
- blend – 5

It is interesting to see that the most frequent category is the same – misselection. The second least frequent class of errors is overinclusion, which corresponds to Stemberger's *Additions* – the last on the list. As for the reasons why *misordering* errors are absent from Stemberger's list, it seems that this category is not relevant to NS slips – i.e. misorderings did not occur in the subjects' language production. That was actually the key finding of the study, whose aim was to question the psycholinguistic validity of the concept of deep structure and transformations, and to give support to alternative syntactic theories. The very low frequency of blends in the data may result from the fact that the analysis is limited to transfer errors, which must be rare in that category.

1.6.1.3. Other descriptive classification schemes

Apart from the two major taxonomies presented above, there are numerous other systems of classifying errors descriptively. They include the following distinctions:

Covert vs. overt errors

The distinction introduced by Corder ([1971] 1981) accounts for the fact that learners may produce language forms which are evidently deviant (overt errors), or forms which are superficially well formed, but still do not express the meaning the learner intended to express or that result from idiosyncratic rules (covert errors). The latter are much more difficult for the researcher to recognize. Only a lot of experience in teaching particular L1 learners and knowledge of that L1 makes it at all possible to realize.

Local vs. global errors

The distinction between local and global errors was made by Burt and Kiparsky (1974) and is defined as follows:

> Global mistakes are those that violate rules involving the overall structure of a sentence, the relations among constituent clauses, or, in a simple sentence, the relations among major constituents. Local mistakes cause trouble in a particular constituent, or in a clause of a complex sentence. These are relative notions; something that is global in one sentence may become local when that sentence is embedded in a bigger sentence. (Burt and Kiparsky 1974:73)

Global errors are more important from the point of view of language teaching and learning, as according to Burt and Kiparsky they are more likely to prevent successful communication from taking place.

Domain vs. extent

Instead of using a traditional linguistic taxonomy, descriptive classifications may define the *level* of language system affected by the error. Hence, Lennon (1991) defined *error domain* as "the rank of the linguistic unit which must be taken as context in order for the error to become apparent" (Lennon 1991: 191). The levels include morpheme, phrase, sentence and larger units of discourse. Lennon also admits the level of the phoneme, if phonological error were to be analyzed. Error *extent* is "the rank of the linguistic unit,

from minimally the morpheme to maximally the sentence, which would have to be deleted, replaced, reordered, or supplied in order to repair production" (Lennon 1991: 191). Again, the phoneme was also included as a possible unit.

Type vs. token

The concepts of type and token originate from corpus linguistics, where they were introduced for quantitative purposes. "[F]or two errors to be distinct types, they must, in general, be distinct at the level of lexical realization. If two errors are lexical replicas they are regarded as tokens of a single error type" (Lennon 1991:186). Such a system allows researchers to measure the frequency of an error involving a given lemma in the body of data they analyze, in comparison, for example, to other errors.

Straightforward as this distinction seems, there may still be some problems with applying it to real-life data. For example, among the errors collected for this study instances of the word **psychologic* occur. The question arises whether the error type should be identified here as the lexical item *psychological* which has not been acquired properly, or perhaps it should be broadened to the use and misuse of the adjectival word formation suffixes {-al} and {-ic}. The latter would be much more informative for pedagogical purposes, indicating a need to spend more time on and pay more attention to this particular aspect of word formation, especially that there are numerous other errors that could be recognized as tokens of the same type (e.g. *economical* instead of *economic*, or *magic* vs. *magical*).

The task criterion

Errors may be classified on the basis of what type of language performance task the learner was engaged in when the error occurred. The distinctions may be very broad (e.g. speaking, writing, translation) or quite narrow; for the sake of comparative analysis of IL variation dependent on tasks, some studies used narrowly defined task types: free oral production, elicited imitation, written sentence combining and grammatical judgment (Tarone 1983), or written grammaticality judgment, oral description task, oral interview and oral narrative (Tarone 1985).

***Systemic errors* ('błędy systemowe') vs. *language norm errors* ('błędy poprawnościowe')** (Szulc 1984)

The distinction between systemic errors and language norm errors is connected with two types of criteria applied to error recognition discussed above: grammaticality and correctness respectively. The distinction is parallel to the two main approaches to language in general: descriptive and prescriptive. The former is preferred in linguistics and is an objective observation of what forms and structures native speakers of a particular language variety use and conceive of as part of their language. The latter is more common in conservative educational contexts, where either teachers or language learning materials make recommendations to students, indicating what language forms are associated with high social status and are considered *correct*, and which forms are seen as non-standard and/or socially marked. If we move this distinction into the field of error analysis, a *systemic error* would be a language form or its use which does not occur in the language system of a particular variety as it is used by its native speakers and so can be recognized by them as 'foreign'. *Language norm errors*, on the other hand, would be elements which do appear in native speakers' use of the language in particular contexts, usually either very informal or local, but which are not accepted as part of the standard by the educated public or, as is the case in some countries, by such language regulating bodies as the *Académie française* ('The French Academy') or *Rada Języka Polskiego* ('The Council for the Polish Language'). Although controversial, such a distinction is quite valid in the EFL context, where it is important to define standards and develop in learners an awareness of the social norms of the target language.

1.6.2. Explanatory taxonomies

Apart from *describing* the IL forms the learner has produced, error analysis often attempts to *explain* them, that is, account for the sources of the errors and try to recognize the processes that led to their occurrence. Therefore, another group of error classifications must be added.

1.6.2.1. Language system

A classification based on the recognition of the *language system* which lies at the source of the error (Richards 1971) includes:

- *interference errors* (also called *interlingual* or *transfer errors*): The error is a result of the learner's L1 system features being transferred to IL, where knowledge of L2 is insufficient to express the intended meaning;
- *intralingual errors*: These are committed as a result of misconceptions about L2 rules, wrong application of actual L2 rules, and the mutual interference of items within the target language;
- *developmental errors*: Developmental errors are similar to intralingual ones, except that they are identical with errors committed by children acquiring a given language as their L1. Originally, this category was included by Richards (1971), but is not usually identified as separate from intralingual errors by those researchers who follow the distinction (e.g. Schachter and Celce-Murcia 1977).

1.6.2.2. Strategies

Errors can be classified depending on what *strategies* applied by the learner caused them (James 1998). The categories here include:

- *learning strategy-based errors*: false analogy, misanalysis, incomplete rule application, exploiting redundancy, overlooking co-occurrence restrictions, hypercorrection (monitor overuse), overgeneralization or system-simplification;
- *communication strategy-based errors*, which include:
    - holistic strategies (approximation), in which the learner seeks a synonym, superordinate, or antonym of the intended word, or coins a new word;
    - analytic strategies (circumlocution), by means of which learners try to communicate elements of meaning of the word they need (description, narration, a list of associated words);
- *natural* vs. *induced errors*: the distinction made by Stenson (1974) based on the observation that some errors occur as a result of *natural* code-breaking strategies

adopted by the learner, while others are *induced* by the language instruction provided by the teacher in class;

- *look-up errors* (James 1998: 199): errors resulting from indiscriminate use of dictionaries (especially bilingual ones) and other reference materials;
- Corder's ([1974] 1981) *systematicity levels*, according to which errors can be:
    - pre-systematic, where the learner is unaware of a TL rule;
    - systematic, where the learner arrived at a rule which is wrong;
    - post-systematic, where the learner knows the rule, but uses it inconsistently;
- *error* vs. *mistake*: This distinction made by Corder ([1967] 1981) stresses the difference of the significance between systematic *errors* of competence, which can provide a lot of information about the learner's IL, and unsystematic *mistakes* of performance, understood to be slips of the tongue (or pen), possible in L1 as well, and of no significance to foreign/second language learning. In practical terms this distinction is very difficult to sustain, especially in studies involving analyses of large amounts of language material.

Conclusions

Errors are central to the understanding of language acquisition processes, and the way in which teachers respond to them can have a strong effect on their students' success. As it has been pointed out in the chapter, both learners and teachers appreciate the importance of corrective feedback. All theories within the field of applied linguistics offer their recommendations concerning errors and how they should be addressed. Admittedly, error analysis, which is closely connected with school reality, still evokes negative connotations in some researchers because of its connections with behaviorism. Attitudes are slowly changing, though: the development of corpus instruments has opened new possibilities in diagnosing learners' problems, not only in research but also in day-to-day classroom teaching. Thanks to this, applied linguists' interest in learner errors seems to have revived.

Interference errors are a special case for both researchers and teachers. The former find them particularly controversial because their L1 origin puts them at the core of

the debate concerning the very nature of second/foreign language acquisition. For behaviorists, transfer is simply ‘habit’, to be eliminated where it interferes with learning; for Krashen it is nothing more than ‘padding’, one of many communication strategies; for Schachter (1992) it constitutes a constraint on learners’ hypothesis testing; and for Kohn (1986) and Odlin (1992) it is a process in which learners employ their knowledge of L1 in processing L2 input for both learning and production. Access to universal grammar in L1 acquisition as compared with L2 acquisition is another contentious issue among linguists, which affects the way in which the role of L1 is understood. One of the most important issues connected with cross-linguistic influence, however, is its significant role in fossilization. If, as Selinker and Lakshmanan (1992) claim, negative transfer is a sufficient or even necessary condition for fossilization to occur, then care needs to be taken in the classroom to minimize its influence. This is particularly difficult in monolingual classes, where learners are likely to reinforce one another’s transfer errors and make them permanent, if the balance between communicative activities and native-speaker input is not preserved.

The effects of transfer can vary, depending on two groups of factors: structural (connected with various qualities of the languages involved) and non-structural (related to individual characteristics of the speaker, e.g. age, ethnic background or level of proficiency). Knowing the students’ situation, the teacher can predict whether interference could be a strong negative factor in their learning, and provide the relevant support in overcoming it. There are a variety of instruments that can be chosen for the purpose, and data-driven learning techniques are among those well worth considering.

Chapter 2: The theoretical background for data-driven learning

Introduction

This chapter discusses the connections between data-driven learning, theories of conceptualization, learning theories, and language pedagogy. DDL finds significant grounding in general theories of learning, whose ideas originate with ancient philosophers like Socrates or Aristotle, or more modern ones like Locke and Kant, and their considerations of how human knowledge is formed (cf. Siek-Piskozub 2006). The underlying premise here is that knowledge is constructed by means of understanding gained from experience, that the key element in its development is the individual, Kant's "knowing subject". This stands in opposition to Plato's theory of ideas, or, a more recent notion, Popper's concept of "the third world", a body of objective "contents of thought" and "theoretical systems", which exist independently of human psychological reality (Popper 1968: 333f). One could speculate, for example, that "the third world" could hold Chomsky's *competence*, the ideal concept of language, among its theoretical systems. DDL is grounded in the opposite view, most expressly represented by constructivism: knowledge is built by the learner as s/he is involved in the active process of structuring it from the available information around them. These ideas are discussed in the current chapter, and are supplemented with views from another learning theory, i.e. connectionism. Next, those trends in language pedagogy which share certain elements with DDL are reviewed: the Natural Approach, the Lexical Approach and Form-Focused Instruction. The chapter begins, however, with an outline of the key notions of corpus linguistics, followed by a presentation of its instruments. Thus the reader is pro-

vided with some background information about the field within which DDL was developed.

2.1. Language corpora – definition and types

A language corpus is very generally defined as "a collection of texts, written or spoken, which is stored on a computer" (O'Keeffe et al. 2007: 1). Although "analog" collections of texts are sometimes referred to as corpora as well (e.g. in literary studies), the digital format of the corpus is crucial to enabling the kind of analysis that is performed in corpus linguistics; most quantitative and statistical analyses would be impossible to perform if the data were stored in the conventional, printed format. Corpora vary greatly in terms of size, genre, register, mode (written/spoken), language variety (standard/social dialect/geographical dialect), and numerous other features, all depending on the principle according to which texts were selected and accumulated. A slightly more precise definition of a corpus is offered by McEnery and Wilson (1996: 177), who define it as "a finite collection of machine-readable texts, sampled to be maximally representative of a language or variety". Following is a list of different classifications of corpora, used in corpus linguistics and language pedagogy (based on Gabrielatos 2005, Waliński 2005, and Krajka 2007b). The classes, which may overlap and which are by no means mutually exclusive, are:

- *representative*/*reference* corpora (LOB, the Brown Corpus, the BNC at Lancaster University) – large corpora with a balanced representation of various text types (genre, content, similar text length) vs. *monitor* corpora – corpora which are systematically updated and which keep expanding with time; the main basis of reliability of the latter is their size (e.g. the Bank of English at Birmingham University, COCA - the Corpus of Contemporary American English);
- *sample* corpora – containing samples of texts of similar length (LOB – about 2,000 words for each source text) vs. *full-text* corpora – containing texts in their entirety;
- *large* corpora (like the BNC or the Bank of English, with about 100m words) vs. *small scale* projects that are usually built by individual researchers for the needs of their own research (several thousand words);

- *general* corpora (with a representation of a variety of general texts on different topics and of different genres, but with few or no specialist texts) vs. *specialized* corpora (with texts representing a given domain or field of study);
- *written* vs. *spoken* corpora – the latter usually collected as transcripts (e.g. the spoken section of the BNC), but sometimes with access to audio recordings as well, e.g. SCRIBE (see Huckvale 2004); many corpora contain both written and spoken data;
- *national* vs. *regional* corpora – the former containing samples from speakers of different regional and social varieties of a given language, and the latter representing a given local variety;
- *synchronic* (representing a language at a given point in time) vs. *diachronic* corpora (text samples from different points in history of a given language community);
- *monolingual* corpora – one language represented vs. *multilingual* corpora – many languages represented; the latter can be further divided into *comparable* corpora – built with the use of the same sampling techniques and criteria for each language represented in the corpus (e.g. subject matter, style, medium, time) and *parallel* corpora – with *the same* texts (source texts and their translations) in two or more languages;
- *plain text* (orthographic) corpora – those which are composed of simple text files vs. *annotated* (tagged) corpora – corpora which contain POS annotation, lemmatization, parsing, semantic information and/or metatextual information;
- *native* speaker corpora vs. *learner* corpora – the former containing texts produced by native speakers of a language, the latter documenting how it is produced by people who learn it as a second or foreign language (e.g. ICLE – the International Corpus of Learner English or PICLE – its Polish component compiled at Adam Mickiewicz University, see Kaszubski 1999, 2002 and 2004).

The choice of a corpus for a given study depends on the kinds of questions the study poses; if any generalizations are made on the basis of the data, they should take account of the types of discourse represented in the corpus, the general principle of its composition as well as its size.

2.2. Key notions of corpus linguistics

The different types of corpora presented above can be utilized in numerous ways within the field of corpus linguistics. Biber et al. (1998: 157) define the corpus-based approach to language as one "using a large, principled collection of texts and combining quantitative and computational techniques with qualitative interpretations".

Conrad (1999: 3-4) discusses three major defining features of corpora and the effect they exert on research outcomes. These are:

- the use of *principled* collections of naturally-occurring texts, i.e. corpora;
- the use of *computers* for analysis;
- the inclusion of both *quantitative* analyses and *functional* interpretations of language use.

The use of corpora allows for stylistic and other genre-related generalizations and comparisons to be made. As discussed above, corpora can have various defining principles, and so generalizations and observations based on them must take those principles into consideration. Comparisons between the results of studies based on differently composed corpora, as well as other quantitative analyses of them, need to take those principles into account. As for the use of computers for analysis, it opens new possibilities to researchers by giving access to information retrieved in automatic or semi-automatic processes. Not only rich statistical text data but also the variety of ways in which the concordances can be arranged and easily rearranged would be unobtainable by means of manual, line-by-line analysis. Such options facilitate linguistic observation and allow researchers to discover new facts about language. Finally, combining quantitative analyses with functional interpretations of language use permits linguists to discern a variety of patterns that would not otherwise emerge. The ability to combine statistical data with determining the possible conditioning factors for various lexicogrammatical choices gives corpus analysis special value: "[T]he more qualitative, functional interpretations describe the communicative functions that correspond to the quantitative patterns that we find" (Conrad 1999: 4).

The theoretical foundations of corpus-based language analysis were laid by Sinclair (1991), who was primarily interested in language as a social phenomenon, an instrument of communication among real people, rather than as an abstract system whose analysis was based on the intuitions of individual researchers. He wanted it to be ana-

lyzed empirically, on real-life data now available electronically. Intuitions about language are valuable in that they show how people conceptualize language, but that is not the reflection of how language really works. Actual language use is seen as the best evidence available for recognizing the ways in which language operates as a system. Sinclair (1991: 6) summarized this opposition very accurately in the following terms: "One does not study all of botany by making artificial flowers".

The considerations above reflect the major division in linguistics between the rationalist and empiricist approaches to language. The former is founded on the concept of language competence as understood by Chomsky:

> Linguistic theory is interested primarily with an ideal speaker-listener, in a completely homogenous speech community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance (1965: 3).

The latter analyzes actual language data, produced by real people in their various social, geographical and historical contexts. One of the most recent areas of study within this approach to language is Natural Language Processing (NLP), which accounts for the actual use of language and depends on statistical analyses of spoken and/or written language data, usually obtained from various types of corpora (Manning and Schütze 2001: 4). Understandably, the latter approach is more willingly embraced by applied linguistics as it is relevant to many current research areas within its scope, such as machine translation, information retrieval, speech synthesis, forensic linguistics and language pedagogy.

Sinclair's view of language is in stark opposition to Chomsky's (1965) generative, rationalist models, according to which language constitutes a separate component of the human mind, an abstract system of principles which can be accessed by the native speaker's intuition (*competence*) and is only indirectly reflected in textual data (*performance*). Generativists see the well-formedness of an expression in categorical terms: a sentence can be either *grammatical* or *ungrammatical*, depending on whether it conforms to the internal structural rules of language. Appropriacy and naturalness, the assessment of which would require reference to text-based frequency data or at least consideration of the situational context, are left out of the equation altogether.

The very concept of grammar in corpus linguistics is unique, and is tied in very strongly with Sinclair's overall approach to language. Sinclair (1991) postulated that two different and complementary principles were needed for the interpretation of language text. These are:

- open choice principle:

> This is a way of seeing language text as the result of a very large number of complex choices. (…) It is often called a 'slot-and-filler' model, envisaging texts as a series of slots which have to be filled from a lexicon which satisfies local restraints. At each slot, virtually any word can occur (Sinclair 1991: 109).

- idiom principle:

> There are sets of linguistic choices which come under the heading of register, and which can be seen as large scale conditioning choices. Once a register choice is made, (…) then all the slot-by-slot choices are massively reduced in scope or even, in some cases, pre-empted. (…) The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments (Sinclair 1991: 110).

In other words, grammatical choices which result from general rules of language structure are strongly limited by the "language chunks" that speakers choose as they construct utterances. These chunks, previously called *prefabrications* or *prefabs* by Bolinger (1976), are whole lexical elements which constitute single choices in language use and are stored in the memory as such, along with individual words. According to Partington (1998), the use of prefabs allows much faster processing on the part of both the speaker and the hearer. The number of choices that need to be made in processing language is then reduced, and fluency in both the production and the perception of language is facilitated, especially in synchronous communication (face-to-face conversation, telephone conversation or an online chat, for example). As far as asynchronous communication is concerned (literary texts, business correspondence or pre-recorded speeches, for example), chunks also have a very important role to play: they are strong indicators of style and register, allowing the reader to recognize the genre of a given text and thus process it more easily. Legal, academic or business English, for example, all have their unique lexical bundles by which they can be recognized. Appropriate use of idiomatic phrases, collocations and other prefabricated lexical elements makes the impression of fluency and confidence. That is why it is crucial for advanced language

learners to master them and learn vocabulary in ways that make remembering clusters easier. Concordance-based activities definitely do just that.

While the open-choice principle can be recognized in other linguistic theories (e.g. structuralism or generative grammar), the idiom principle is more strongly associated with corpus linguistics. Both principles are, however, fundamental in the corpus approach to language study, and are exploited in the very idea of a concordance, its key instrument (see section 2.3.). Although separate and mutually exclusive, the open-choice and idiom principles are very difficult to distinguish in their actual implementations in real texts because transitions from one to the other are not signaled by any overt markers. It is challenging, if not impossible, for a researcher to decide whether a particular form was generated through the open choice principle or the idiom principle. Therefore, according to Sinclair (1991: 104) "it is much more fruitful to start by supposing that lexical and syntactic choices correlate, than that they vary independently of each other". Hence he came to the conclusion that the division between grammar and lexis is artificial and inadequate: grammar is mostly word-specific and cannot be considered in abstraction from lexical choices made by the user. A word has its local, unique grammatical features, which are different for its every meaning. Those features include such aspects of grammar as countability, complementation, agreement, or choice of tense and aspect. This is why Sinclair postulated joining lexis and grammar into one category: lexicogrammar. He was not the first to do this, though; lexicogrammar was also strongly endorsed by Halliday (1985: xiv), who went so far as to say that "syntax and vocabulary are part of the same level in the code". He understood it to be a functional tool of the "meaning potential" – an instrument which speakers have at their disposal to encode their meanings. Lexicogrammar is therefore a very broad concept which prevents linguists from imposing arbitrary and rigid distinctions on the language phenomena they examine, and which makes their analyses more consistent.

A majority of corpus-based studies employ the concept of lexicogrammar, most often to include all word- or phrase- and sentence-level aspects of the material analyzed (as is the case in this project), and sometimes as an in-between category, *apart from* grammar and lexis, which are still preserved as categories in their own right (e.g. Dagneaux et al. 1998). It seems that the latter solution is not in agreement with Sinclair's original intentions behind the concept of lexicogrammar and adds to the confu-

sion instead of limiting it. In other words, the concept of *lexicogrammar* is a way of avoiding arbitrariness in classifying language problems; it offers a wide category which encompasses overlapping and co-occurring rules involving the use of words. A good example of such classification problems is the word *suggest* and its colligation, i.e. the range of forms with which it can appear (gerund, *that*-clause with subjunctive or *should* + infinitive, but not a noun, pronoun or *to*-infinitive). The rules clearly involve grammar, but are tightly connected with a single lexical choice. Classifying them as either grammatical or lexical problems would always be arbitrary, so it is more convenient to have one broad category of lexicogrammar in which to place them.

Corpus linguistics emerged, then, from a usage-based model of language. Langacker (1987: 66) saw *usage event* to be the key to understanding language and defined it as "a symbolic expression assembled by a speaker in a particular set of circumstances for a particular purpose". In other words, the speaker chooses the most suitable form of expression for his/her meaning in a given situation. Grammar (or what Langacker refers to as "linguistic convention") is merely a set of symbolic resources that the speaker employs for the target structure to achieve his/her aims in that particular situation. It provides symbolic units and patterns of higher complexity which are *conventional*; the speaker's choices in a usage event can then be *sanctioned* by grammar if they agree with the conventions. The emergence of such a grammar is an effect of the processing of instances of language as they appear in various situations:

> Instances of language are language forms used in a context with a particular meaning. The processing of instances includes the production and comprehension of language (form and meaning) within a particular context, and this processing of instances leads to the formation of more abstract categories that make up the stored representations known as grammar. The grammar created is a network of form-meaning pairs with lexically-instantiated form-meaning pairings linked with categories at different levels of abstraction over these lexical instances (Barlow 2011: 22).

In such a view of language, well-formedness is a matter of degree and individual judgment of how closely an utterance corresponds to linguistic convention, which can only be recognized in reference to what forms speakers actually use in particular situations. It is not a categorical, binary statement. Thus Langacker (1987) abandoned the notion of the *grammaticality* of a language expression in favor of the concept of *degree of conventionality*, which is by nature dynamic and depends on the attitude within the language community towards a given form or expression. Grammar is then defined as

"a structured inventory of conventional linguistic units" (Langacker 1987: 57) and as such is always dynamic and evolves constantly, in tune with general language use. This, on the other hand, ties in with other theories of language, conceptualization and cognition, which are going to be discussed later in this chapter. The concept of the dynamic nature of grammar sheds a new light on the reliability of the key instrument in linguistics: grammaticality judgments. This issue is thoroughly discussed by N. Ellis (2002), who questions the validity of such judgments, saying that they are not only inconsistent among various speakers but are also unstable for the same individual over time. According to the results of research on priming (Luka and Barsalou 1998, as cited by N. Ellis 2002: 162), perceptions of grammaticality are affected by a speaker's recent exposure to particular structures. Ellis concludes:

> These standard, implicit learning and memory effects for grammatical constructions show that competence as assessed using grammaticality judgments seems hardly more constant than grammatical performance. Both are affected by frequency and recency of use of construction. Grammatical analyses that a language user has frequently or recently experienced are preferred to analyses that must be newly constructed (2002: 162-163).

Since grammaticality judgments are not as dependable as Chomsky presumed, linguists turn more and more willingly to language corpora as reference, because they make frequency of occurrence available as at least one of many criteria of the assessment of grammaticality or acceptability.

2.3. Corpus linguistics – techniques of analysis

Corpus linguistics has developed its own instruments of inquiry, which include concordancing, word frequency counts and lists (ordered alphabetically, by frequency or by first occurrence), key word analysis, cluster analysis, and lexicogrammatical profiles (O'Keefe et al. 2007). These may be applied to a single text, a corpus of the same type of texts, or a collection of various texts representative of a language or variety. Depending on the type of corpus and the questions asked, the data obtained through these instruments can be utilized in different ways.

Concordancing is the most essential instrument in corpus linguistics. Its outcome presents a key word or phrase in its immediate context, placing examples of its use line

by line, usually with options as to the context alignment (often referred to as KWIC – Key-Word-In-Context). The search word is often called "node", and is usually presented in the very middle of each line. It is this arrangement of data that allows researchers and learners to observe prominent features of the usage of the node. The KWIC format makes patterns easy to recognize and to remember. A concordance shows a number of examples, which offer information about the possible differences in use and meaning to researchers and enhanced input to learners. Figure 6 shows an example of concordance lines for the verb *raise* (in its base form). A brief analysis of the concordance reveals the verb's major meanings (both literal and metaphorical), collocations and patterns of use (a transitive structure):

should ask first whether anyone wishes to	**raise**	any questions arising outof those er,
from Saturday to Monday, is being held to	**raise**	cash for the church repair fund.
badges and other ceremonies needed to	**raise**	company consciousness? The presence of
fairly light, but from time to time can	**raise**	dust . Now and then a fine drizzle of
competencies, benchmarks are intended to	**raise**	expectations . What happens in practice
suffering from cancer. They also hope to	**raise**	funds for research into brain tumours.
1950s when the local authorities began to	**raise**	money for capital projects by issuing
She paid tribute to all who helped	**raise**	money for the appeal which now stood at
creation of Education 2000, a trust to	**raise**	money from industry. It would be just a
see the problem solved, only as ever to	**raise**	more. Lecturers like these could
take the title as it stood, and would not	**raise**	requisitions after contracts were
but then I thought I might as well try and	**raise**	some money for the hospice which is on
in Latin America and Africa. Pay-offs	**raise**	the cost of development projects and,
Nevertheless, Clara had been obliged to	**raise**	the subject of a best dress, because
was accepted by the Board er asking us to	**raise**	this matter with the the Roman Catholic
CHIN ELIMINATOR. Lying on your tummy,	**raise**	your head and shoulders. Point the chin
please me." "Oh get out, go home!" "Do n't	**raise**	your voice. I know your nerves are on

Figure 6. Sample concordance lines for *raise* from the BNC

Word frequency lists can offer an insight into some features of the text or set of texts analyzed, especially if compared with another type of text. The relative position of personal pronouns at the top of the list, for example, could reveal whether the texts in the corpus tend towards personalized, interactive presentation (first and second person) or a more objective and detached point of view (third person). This is very well evidenced by the following comparison of the ten most frequently occurring words in seven different corpora:

Table 3. Comparison of word frequencies for the ten most frequent words across six different datasets (adapted from O'Keefe et al. 2007: 11)

	1	2	3	4	5	6	7
Rank order- der	Shop (LCIE)	Friends (LCIE)	Academic (LIBEL)	Australian Corpus of English	CIC news- paper and magazine sub-corpus	BNC sub- corpus[15]	KJO blog corpus
	spoken	spoken	spoken	written	written	written	written
1	you	I	the	the	the	the	I
2	of	and	and	of	to	of	the
3	is	the	of	and	of	and	to
4	thanks	to	you	to	a	to	and
5	it	was	to	a	and	a	a
6	I	you	a	in	in	in	of
7	please	it	that	is	is	it	my
8	the	like	in	for	for	is	it
9	yeah	that	it	that	it	was	in
10	now	he	is	was	that	I	is

The first six corpora listed in the table above are different sections of the Limerick Corpus of Irish English (LCIE), the Limerick-Belfast Corpus of Academic Spoken English (LIBEL), the Australian Corpus of English (also known as the Macquarie Corpus), the Cambridge International Corpus (CIC), and the British National Corpus (BNC). The seventh column provides data for the KJO blog corpus constructed for the needs of this study. They can be characterized as follows (cf. O'Keefe et al. 2007:11):

(1) shop encounters (8,500 words);
(2) female friends chatting (40,000 words);
(3) spoken academic English (one million words);
(4) written Australian English (one million words);
(5) written British and American English (one million words);
(6) written section of the British National Corpus (ca. 90 million words)
(7) written learner English from blog postings (L1 – Polish, ca. 224,000 words).

The comparison reveals some stylistic features of the texts in each of the corpora, and reflects their function: the KJO corpus, for example, although written, shares some items (and their relative positions) with the 'friends' spoken corpus and some with the

[15] The data in this column have been added to the original table on the basis of the information provided on the official BNC website at http://ucrel.lancs.ac.uk/bncfreq/lists/2_3_writtenspoken.txt

written ones. The top rank of the first person singular pronoun indicates high speaker/ author involvement and a subjective point of view. The absence of the pronoun *you* from the top ten, however, indicates the less interactive character of the texts there than those in any of the spoken corpora. The 'friends' corpus, on the other hand, is distinct from all the others by the high position of the word-form *like*, a common conversational filler word, which would not be used even in very informal writing. There is much more that can be read from Table 3, but even this short analysis offers an interesting sample of what kinds of information frequency lists can provide.

Another function employed in corpus studies is key word analysis. It isolates words which are unusually frequent in the text in comparison with the predefined 'norm', usually based on a representative corpus. In this way lexicogrammatical features of various text genres or registers can be studied (cf. Biber 1993 or Conrad 1999, for example). The process requires the data to be organized into frequency lists – for both the text and the reference corpus. Any word which is in some way outstanding in its frequency in the text is considered to be "key" (Scott 2012). The outcome of the process is presented in the form of a list of those words whose frequency is extraordinarily high in comparison to the norm (*positively key words*) or unusually low (*negatively key words*). Such lists are usually ranked by frequency. It is these words that are unique to that particular text, and a list like this reveals the subject matter of the text as well as its lexical density.[16]

O'Keefe et al. (2007) list one more mode of analysis in corpus linguistics: cluster analysis, a concept that features very strongly in corpus analysis owing to its immediate connection with Sinclair's idiom principle discussed at the beginning of this chapter. Clusters, often referred to as *chunks*, are groups of word forms that tend to occur in linear sequence, e.g. *a couple of* or *that sort of thing*. This distinguishes them from phrases and collocations, which are more flexible and can have some 'open' slots in the sequence: *the dog barked*, *the barking dog*, *the barking of a dog*, etc. There is a connection between the key words in these phrases, but they are not absolutely fixed, as *lexical bundles* are (another term for clusters, from Biber et al. 1999). Cluster analysis performed with corpus tools is relatively straightforward: the text analysis software can

[16] Such functions are included in more advanced text analysis software, e.g. ConCapp (Edict 2011), Wordsmith (Scott 2012), AntWordProfiler (Anthony 2012), or their online equivalents such as KeyWords (Cobb 2011).

generate a list of words occurring together in the corpus, the length of the string having been previously defined by the user. Thus lists of two-, three-, or four-word strings can be produced, and arranged either by frequency or alphabetically. Such lists provide useful information about the text or texts in the corpus, their style, register and other unique features of language. Cluster analysis can be useful in interlanguage studies as well: "[T]he use (or non-use) of lexical bundles by second-language learners has been considered a useful yardstick for the comparison and evaluation of learner competence vis-à-vis native speaker competence" (O'Keefe et al. 2007: 62). For such comparisons to be accurate, the corpora would have to be very well matched in terms of mode (written/spoken), genre(s), and author characteristics, especially if longer strings were to be analyzed. Short strings (two or three words long) are less content-dependent and are mostly composed of function words. Table 4 below shows the top 20 three-word strings from two native speaker corpora and the KJO learner blog corpus. The unique character of blog writing is reflected in the composition of the last list, as it includes some typically written and spoken language phrases. Most of them include a first or second person pronoun, indicating a monologic nature of the texts that makes them similar to spoken language in some ways. The number of phrases identical to those from the native speaker corpora is not impressive, but considering that what is being looked at here is a very small sample of L2 English if compared to the other corpora, this is hardly surprising. What is more important is that all of the phrases are common expressions in standard English. The problem that could be posed here is whether perhaps some of them could be overrepresented or underrepresented in the corpus as compared to native English frequency of use. The corpus gathered for this study may not be big enough to allow such generalizations. What is more, the topics addressed in the texts are determined to a large extent by the teacher's assignments, which is why some clusters emerge on the list with higher than 'normal' frequency (e.g. *when I was* must be particularly frequent because of the task which required students to post their childhood photographs and comment upon them). Another string of words that is on the L2 list, *the fact that*, seems a more likely case of overuse; the author has observed that her learners use it very often when they have difficulty building *wh*-clauses. This observation seems justified once the frequencies are considered: the frequency of the phrase in the blog corpus is 41.1 per 100,000 words, while in the written section of the BNC it is only 13.5 per 100,000 words (BNCweb 2008). This disproportion is too big to result from differences between

the corpora, especially considering that the phrase in question is neutral in terms of topic. Also, in Scheffler (2008: 238) the phrase is ranked fourth on the list of the most frequently overused three-word clusters in the corpus of university level argumentative essays by Polish students of English. If the relatively relaxed style of blog writing is considered, the phrase is definitely overrepresented in the KJO corpus.

Table 4. Lists of the top 20 three-word clusters from different corpora
(based in part on O'Keefe at al. 2007: 66-68)

	CANCODE spoken corpus (5m)		CIC written corpus (2m)		KJO blog corpus (ca 214 k)	
	item	frequency	item	frequency	item	frequency
1.	*I don't know*[17]	5,308	*one of the*	1,886	when I was	120
2.	*a lot of*	2,872	out of the	1,345	*a lot of*	119
3.	I mean I	2,186	*it was a*	1,126	*one of the*	111
4.	I don't think	2,174	there was a	1,083	I have to	108
5.	do you think	1,511	the end of	1,045	would like to	103
6.	do you want	1,426	*a lot of*	785	I want to	100
7.	one of the	1,332	there was no	753	I would like	97
8.	you have to	1,300	as well as	737	the fact that	91
9.	*it was a*	1,273	end of the	691	*to be a*	90
10.	you know I	1,231	*to be a*	672	*I don't know*	89
11.	you want to	1,230	it would be	671	I decided to	72
12.	you know what	1,212	in front of	655	you can see	68
13.	do you know	1,203	it was the	643	one of my	65
14.	a bit of	1,201	some of the	621	I think that	60
15.	I think it's	1,189	*I don't know*	604	*it was a*	60
16.	but I mean	1,163	on to the	602	I wanted to	59
17.	and it was	1,148	part of the	600	I was a	59
18.	a couple of	1,136	be able to	596	this is the	54
19.	you know the	1,079	the rest of	577	as far as	53
20.	what do you	1,065	the first time	567	I had to	52

Even without making comparisons, however, cluster lists can provide valuable information. Table 5 below presents a list of the most frequent four-word clusters in the blog corpus:

[17] The phrases in italics are the ones that appear both in the blog corpus and in either of the native speaker corpora.

Table 5. List of the top 20 four-word clusters from the KJO blog corpus

I would like to	91
to be a teacher	37
as you can see	34
at the same time	29
when I was a	28
I must say that	27
from time to time	26
one of my favourite	26
one of the most	26
I don't want to	25
for the first time	24
I have to admit	24
when it comes to	24
I am going to	23
in the middle of	23
if you want to	22
at the age of	21
I must admit that	20
of the fact that	20
one of the best	19

The list offers some insight into the style and content of the texts included in the corpus, as well as the level of fluency the students demonstrate. The phrases included in the list are mostly colloquial, and most of them include first and second person pronouns, for reasons discussed above. It is also worth pointing out that all the phrases at the top of the list are acceptable English (unless they are used inappropriately), so there is no evidence of persistent L1-based or other noticeably non-native phrases marking the students' performance. Errors do appear, of course, but they are much further down on the list, where idiosyncratic problems are placed, rather than forms recurring in the students' production. In the few cases where the error is repeated two or three times, it happens to be committed by the same author. In this way errors of form (not of use) at the phrase level can be analyzed, to establish if there are any recurrent problems. In this particular study this did not seem to be the case.

Corpus tools provide information that has always been available from texts, except that they facilitate access to it enormously and make studying language a transformed experience. What could be obtained through laborious line-by-line analysis can now be retrieved within seconds, leaving linguists, teachers and language learners more time and resources for posing new questions and analyzing the results generated through corpus tools. Also, the forms in which the data are retrieved (concordances and

frequency lists) allows researchers to make observations that would be otherwise very difficult, if not impossible, to arrive at.

2.4. Grammar and conceptualization in corpus linguistics

Corpus linguistics approaches language in new, unprecedented ways. Its development was made possible thanks to advanced technology, but its view of language is rooted in ideas which were developed long before modern corpora even existed. On the other hand, corpus-based language study has affected other, more theoretical areas of linguistics and has inspired new trends within it.

2.4.1. Prototype theory

In her seminal paper Rosch presents a series of experiments which proved that semantic categories had internal structure in the sense that some entities were consistently recognized as 'better' examples of a given category than others: "cognitive representations of categories appeared to be more similar to the good examples than the poor examples" (1975: 225). These results undermine the traditional Aristotelian classification of entities into absolute categories "whose membership is defined by an item's possession of a simple set of criterial features, in which all instances possessing the criterial attributes have a full and equal degree of membership" (Rosch 1975: 193). In other words, according to prototype theory our concepts are organized into categories around their most typical representatives, and the boundaries between them can sometimes be fuzzy. Concept formation itself is seen in terms of a statistical procedure: "[A] concept should encode the distribution of statistically prominent properties in a category. (…) [Prototype theory] doesn't aim to monitor whether various properties always co-occur, but only whether they tend to" (Margolis and Laurence 1999: 29). What is more, research confirms that items that are recognized as typical, are learned and categorized faster than those, which are less frequent and unusual (cf. Rosch and Mervis 1975, Murphy 2004).

If applied to the rules of lexicogrammar, prototype theory would indicate that words had some *prototypical* contexts, which were used by the majority of the native

speaker population, and apart from these the words could be used in less predictable ways, marked or idiosyncratic, limited by unique contexts and by special meanings intended by the speaker. Leech suggests that what is known in pedagogical grammar as a "rule of thumb" accounts for 80-90% of (prototypical) cases, while other instances of the word are on the periphery of the general tendency. These should not be seen as exceptions but extensions to the rule, as "grammar typically does not lend itself to clear-cut boundaries" (1994: 20).

From the perspective of cognitive grammar, where grammatical classes are perceived as not principally different from other conceptual categories, Langacker (1987) also postulates that the prototype model could be a suitable approach to analyzing grammar and language constructions. He says that grammatical groupings are prototypical rather than absolute in nature ("the rule/list fallacy" – Langacker 1987: 49). The traditional criterial-attribute model, which assumes full consistency with clearly defined sets of features and no 'fuzzy' areas, cannot account for all the possible ways in which people use natural languages. Cognitive grammar is usage-based, which means that it attempts to account for actual language data rather than to construct an idealized language system, as it is the case with generative grammar:

> The prototype model offers a more realistic account in many instances, but adopting it implies that class membership is not predictable in absolute terms: it is a matter of degree, decreasing as an entity deviates from the prototype, with no specific cutoff point beyond which speakers abruptly become incapable of perceiving similarity and thus assimilating an entity to a category. One would be wrong to claim that the prototype model is non-predictive (…) – but its predictions are statistical rather than absolute (Langacker 1987: 49).

This means that the prototype model does introduce rules, but these rules are not expected to predict all – and only – grammatical constructions as generative grammars attempt to do. Instead, the rules reflect the regularities that occur in real language use and specify the most typical instances of a given structure. Such an approach is very well adjusted to corpus-based language analysis, which offers statistical (frequency) data not only at word level but also at higher levels of analysis (collocations, clusters, patterns of discourse, etc.). The prototype model in conjunction with corpus data provides a new perspective on learner error, its recognition and classification. In many instances the judgment of error is not an absolute distinction but indeed a matter of degree, of how *non-prototypical* a given utterance is for the meaning it was aimed to

express. For example, in the data gathered for the current research project it became apparent that students overused the determiner *some* in front of countable singular nouns. Although not impossible in native English, this use is very rare if compared with the frequency of the indefinite article *a/an*, and is marked. In its strong form *some* is usually emphatically positive (e.g. *This is some office!*), while the weak form carries rather negative connotations, as in *They gave me some book*, for example. Here *some* is not a neutral marker of indefiniteness, as *a* would be, but gives the impression of disappointment and disinterest. Therefore, unless the context indicates that there is indeed justification for the marked use of *some*, its use in front of a singular countable noun should usually be seen as an error. Corpora provide statistical data on language use and constitute perfect reference for such considerations; this is then a good point of departure for consciousness raising with students.

Admittedly, the notion of prototype does introduce an element of uncertainty, or 'fuzziness', into considerations of what is and is not acceptable in learners' language production. Here, errors will not be recognized on the basis of whether a form attempted by the learner is 'possible' or not. It is crucial whether it communicates the speaker's intentions, whether it is recommended in a given linguistic and situational context, and whether it would be natural for a native speaker to use it in such a context. Corpus data can be very useful in these judgments, as they can clearly demonstrate the dominant pattern, and possibly the other choices available. If the form which the learner has used is not the dominant pattern, then a concordance search can verify whether the learner's choice appears in the native speaker data at all, and if so, whether the circumstances are quite common or rather unique. This should help the learner to find out not only what the best choice is but also why this is the case. What is more, this can be done without involving the teacher, which will give the learner the sense of autonomy and empowerment. On the other hand, the teacher, with corpus data available, will be able to support his/her judgments and feedback with appropriate examples and statistics, offering the learner dependable, convincing information rather than dry textbook rules or (non-)native speaker's intuitions.

Another consequence of the prototype approach to language is that it provides strong arguments in favor of selecting course material and designing syllabi on the basis of corpus frequency data, not only in terms of vocabulary selection but also in choosing the most typical phrases and contexts in which words occur. Students should be encour-

aged to use those patterns as the ones which guarantee successful communication, enhance fluency and prevent the occurrence of errors.

Beginners should be presented with the most typical examples, and for them these should perhaps serve as the only point of reference. Instances which contradict these generalized rules should not be brought to learners' attention before they are ready to appreciate subtle differences within the lexicogrammar and build more complex utterances. Their goal is usually to achieve basic level communication with other speakers of the target language rather than appreciate its complexity. For example, the phrase *stop doing something* could be promoted as the model to follow, because it is definitely a much more common and productive pattern than *stop to do something* – a possible but much less frequent form, with limited usability[18]. Only if miscommunication occurs, will a need arise for the meaning of the other form to be discussed. As learners progress in their language development, however, more and more extensions to rules can be introduced, and less typical examples can be included in consciousness raising activities. Advanced learners need to have a broader range of choices at their disposal to make their language precise and appropriate to the situation. They need to know not only whether their choices are comprehensible, possible or grammatically acceptable but also how natural they are. At the advanced level they need to have an awareness of the prototypical use of a form, as well as of what other choices are available and how they affect the way their message is perceived. Developing such awareness can be accelerated and strengthened by learners being exposed to concordances and statistical corpus-based data which will allow them to observe usage patterns and differences in the meaning of superficially similar forms.

It is not the case that whatever fails to conform to the "rule of thumb" should be treated as an error. Learners should be encouraged to use language creatively and flexibly. Usage statistics can, however, be helpful in deciding which forms are 'safe' and which are idiosynchratic and should not really be recommended to learners. Apart from pure statistics, what should be considered in such recommendations is differences in meaning and the style or mode of a text in which a given form typically appears (formal/ informal, written/ spoken). Finally, the fact that a form (or its particular use) does

[18] The difference in frequency of occurrence between the two forms can be observed in the following BNC query: http://corpus.byu.edu/bnc/?c=bnc&q=14112343

not feature in a representative, native speaker corpus at all should be an indication of it being an error, or a dispreferred choice.

2.4.2. Lexical priming

A view of language strongly connected with corpus studies is Hoey's (2005) theory of lexical priming. Even its key notion, collocation, is defined with reference to corpus findings, and in very precise, technical terms:

> [Collocation] is a psychological association between words (rather than lemmas) up to four words apart and is evidenced by their occurrence together in corpora more often than is explicable in terms of random distribution. This definition is intended to pick up on the fact that collocation is a psychological phenomenon, the evidence for which can be found statistically in computer corpora (Hoey 2005: 5).

The theory goes even further in changing the mutual relationship between lexis and grammar than Sinclair (1991) did in his concept of lexicogrammar. While Sinclair raised the status of lexis and merged it together with grammar, Hoey goes even further and gives lexis priority, as a more systematically structured system than grammar. The controlling element in language is collocation, because it is responsible for stringing words together into longer, predefined sequences, which only in later stages of sentence construction are put together by means of grammar. Collocation itself is a result of priming:

> We can only account for collocation if we assume that every word is mentally *primed* for collocational use. As a word is acquired through encounters with it in speech and writing, it becomes cumulatively loaded with contexts and co-texts in which it is encountered, and our knowledge of it includes the fact that it co-occurs with certain other words in certain kinds of context (Hoey 2005: 8).

Priming is not limited to collocation; other co-occurrence rules are also built on this subconscious process: semantic associations, colligations, pragmatic associations, textual relations, etc. The theory accounts for one of the key features of natural language: it sees language as a social phenomenon, spreading throughout a language community in day-to-day communication, with children acquiring lexical sequences through priming. The language system is reproduced through those exchanges, but at the same

time it is being slightly modified with new contexts in which it is used, which explains its dynamic character.

The role of corpora in Hoey's (2005) theory is that of evidence. It is through corpora that word associations can be examined and proved. The theory does not offer any direct recommendations for language teaching, corpus-based or otherwise, but it has been an inspiration for DDL practitioners and adds strong arguments in favor of analyzing language through corpora.

2.4.3. Probabilistic grammars

The development of corpus linguistics has influenced theoretical linguistics, and some of its basic assumptions. Formal models of grammar are categorical in that they recognize a sentence as either grammatical or ungrammatical, without any gradation involved. This is questioned by an alternative approach to grammar often referred to as *probabilistic*, according to which judgments of this kind can be made on statistical grounds (Manning and Schütze 2001).[19] They posit that the probability of a structure is correlated with its degree of grammaticality, and the speaker depends on his/her experience, i.e. amount of exposure to a given structure or form, when faced with a grammaticality judgment. It must be added that probabilistic (or scholastic) grammars depend on the frequency information obtained from language corpora and are crucial in natural language processing (NLP), where they are employed in such applications as machine translation and speech recognition.

The probabilistic approach to language study inspires new research and new theories of grammar. Bresnan and Hay (2008), for example, conducted quantitative, diachronic corpus-based studies of two English speaking communities (New Zealand and US English), in which they observed that the two varieties differ quantitatively in how the animacy of the recipient (indirect object) affects the syntax of *give*. Two alternative structures are available in so called *dative alteration*: "*Who gave you that wonderful watch?* (double object construction) and "*Who gave that wonderful watch to you?*" (prepositional dative) (Bresnan and Hay 2008: 246). The authors analyzed the frequen-

[19] The full name of the model discussed in Manning and Schütze (2001) is Probabilistic Context Free Grammar (PCFG).

cy of either construction in the two varieties and devised a *statistical model* which could very accurately predict which forms would be produced under what circumstances in each of them. The authors of the study claim to have demonstrated that variability in language does not result from the cognitive processes involved in language, or features of performance, but from the variability of input to which speakers were exposed when acquiring language:

> A probabilistic, usage-based approach to grammar is able to account for such variation by assuming that different communities differ in the types and frequencies of the constructions that they are exposed to. However, a probabilistic approach also predicts that variation across space and time should exist in less obvious ways—even affecting the subtle probabilistic choices that are made between two variants which are equally acceptable for that dialect (Bresnan and Hay 2008: 246).

> [T]he variability captured in the statistical model is unlikely to be explained by considerations of 'performance' or cognitive processing resources, since we lack antecedently known differences in cognitive resources between the speaker groups studied. Instead we suggest that the results support statistical theories of linguistic competence—what we have called 'gradient grammar' (Bresnan and Hay 2008: 256).

Gradient grammar, then, shifts focus from what is and is not grammatical to the frequency of the occurrence of a given form (in empirical, corpus-based studies) and the degree to which it is considered grammatical by speakers of different varieties. It can be based on statistical, corpus-based data as in the study discussed above, or on grammaticality judgments. These are rarely categorical and unambiguous, which is why some theoretical linguists have embraced the idea of gradient grammar, too. The concept itself is not new, as it was originally introduced and developed by Bolinger (1961), but it has been gaining linguists' interest with the development of corpora and the massive evidence they provide concerning variety and frequency data for alternative forms in language.

Sorace and Keller (2005) also maintain that the grammaticality judgment of language samples should not be conceived of as a binary choice between "totally acceptable" or "totally unacceptable", but a matter of degree, as different sentences and structures can have varying degrees of acceptability. This stands in opposition to the majority of modern models of theoretical grammar, e.g. Optimality Theory (OT), according to which various forms compete for the status of well-formedness and only one candidate can be selected as the 'winner':

> The heart of the proposal is a means for precisely determining which analysis of an input *best satisfies* – or least violates – a set of conflicting conditions. For most inputs, it will be the case that every possible analysis violates many constraints. The grammar rates all these analyses according to how well they satisfy the whole constraint set and declares any analysis at the top of this list to be *optimal*. Such an analysis is assigned by the grammar as output to that input. The grammatically well-formed structures are exactly those that are optimal in this sense (Prince and Smolensky 2004: 3).

It must be said that OT does not attempt to account for 'real life' data, but, in agreement with the generative tradition, conceives of grammar as an abstract entity.

Some objections have been raised against including probability criteria in theoretical grammars (Crocker and Keller 2006: 240ff.), but discussing them is beyond the scope of this study. The point is that corpus analysis is reaching beyond its normal scope of influence, located within the realm of applied linguistics and usage-based models of grammar, to very theoretical fields of language study.

2.4.4. Exemplar grammar

Another new trend in linguistics which, like gradient grammar, challenges the notion of grammaticality and rule-based grammar is known as the exemplar theory of language. Its main premise is that language categories are built on the basis of memorized tokens, i.e. exemplars (Gahl and Yu 2006: 213): "An exemplar-based speech processing system recognizes inputs and generates outputs by analogical evaluation across a lexicon of distinct memory traces of remembered tokens of speech". According to Bybee (2006), one of the major advocates of the exemplar model, every time a token (i.e. a language item: word, phrase, collocation, etc.) is encountered, it requires categorization, and with every categorization the mental representation of the category itself is slightly affected and modified:

> The major idea behind exemplar theory is that the matching process has an effect on the representations themselves; new tokens of experience are not decoded and then discarded, but rather they impact memory representations. In particular, a token of linguistic experience that is identical to an existing exemplar is mapped onto that exemplar, strengthening it. Tokens that are similar but not identical (...) to existing exemplars are represented as exemplars themselves and are stored near similar exemplars to constitute clusters or categories (Bybee 2006: 716).

Such an understanding of grammar and of its refinement in the process of language acquisition can be associated with data-driven learning and the use of concordances in the classroom. A large number of instances of a particular item, in their immediate textual context, should affect the learner's conceptual structure related to the item in question and, eventually, make his/her performance more natural. One of the fundamental principles of the usage-based approach to language is that "there is an intimate relation between grammatical structures and instances of use: grammar gives rise to usage and usage gives rise to grammar" (Barlow 2011: 21). The effect of reading concordances may not be the same as an encounter with the target item in a communicative situation, which would be a natural hypothesis-testing ground for the learner. However, exposure to the target language feature in the form of a concordance is more intensive than it is in a natural conversation or an authentic or even enhanced text. It should therefore help the learner adjust his/her mental representation of the category and give it a shape that corresponds more closely to its actual use.

Recently, corpus-based studies have become prominent in the field of sociolinguistics, mostly thanks to social indexing, which is often implemented in corpora, in the form of tags providing information about speakers' age, gender, regional origin, employment or other aspects of their social status. An example of such a study is provided by Hay and Bresnan (2006), who examine speech production and perception of pronunciation variants from the point of view of the exemplar theory. A wide range of corpus-based and corpus-driven sociolinguistic research is presented by Baker (2010).

2.5. Learning theories and DDL

The aim of the above overview of corpus linguistics and its theoretical background was to help the reader understand how corpora, the instruments of data-driven learning, are utilized to analyze language and what such an analysis can reveal. This section presents the learning processes involved in corpus-supported learning and in analyzing corpus data. These processes are viewed from both the cognitive and neurophysiological perspectives.

2.5.1. Constructivism – the theoretical foundations of data-driven learning

DDL has an immediate and explicit connection with the work of Piaget and his theories of cognitive development known under the label of *cognitive* (or *psychological*) *constructivism*:

> Fifty years of experience have taught us that knowledge does not result from a mere recording of observations without a structuring activity on the part of the subject. Nor do any a priori or innate cognitive structures exist in man; the functioning of intelligence alone is hereditary and creates structures only through an organization of successive actions performed on objects. Consequently, an epistemology conforming to the data of psychogenesis could be neither empiricist nor preformationist, but could consist only of a constructivism, with a continual elaboration of new operations and structures (Piaget 1980: 23).

Constructivists argue, then, that knowledge cannot arise through mere transmission, where the teacher or parent presents a body of information (e.g. language rules) and the learner passively receives it. Instead, the learner needs to construct his/her own new internal meanings on the basis of experience and the empirical data that have been provided. As Phillips (2000: 7) says, "knowledge is not a mere copy of the external world, nor is knowledge acquired by passive absorption or by simple transference from one person (a teacher) to another (a learner or knower)". A significant consequence of such an understanding of knowledge and learning is that for every learner the image of reality that s/he develops is slightly different. It seems that this is especially true of language learning, where the object of study is governed to a large extent by social convention and undergoes constant change. It is therefore unavoidable that not only non-native but also native speakers of a language understand particular concepts slightly differently:

> [A]lthough the various interpretations are undergirded by a common language that has been socioculturally constructed (…), [the readers] may differ with respect to their understandings of individual words, the range and subtlety of their vocabularies, and the associations that various words have for them. In a sense, then, while language is a human construct, different individuals may construct slightly different things with it, even when they use the same words (Phillips 2000: 4).

Such flexibility can be an asset in a language classroom as well as a complication, depending on the attitudes and individual characteristics of learners. What is important to point out here is that data-driven learning techniques take this into account and allow learners to build their own language constructs, though not without guidance

and information structuring efforts from the teacher. Language input presented in the form of a concordance provides learners with opportunities to make their own observations and generalizations of numerous instances of the use of a language form, thus replicating to some extent the natural process of language acquisition, albeit in a condensed and accelerated manner. The educational context of the activity, however, as well as the instructor's guidance, prompts the learner to make the effort and arrive at a rule, rather than remain a passive observer.

Jean Piaget based his theory of learning on the innate functions of biological, evolutionary mechanisms, which are the same for all humans and do not change with age. "The goal of the functions is construing internal cognitive structures which undergo constant change as a result of cognitive development. These structures have a form of schemata" (Siek-Piskozub 2006: 162). The mechanisms involved are *assimilation*, *accommodation*, *adaptation*, and *organization* and they are referred to as the functional variants of intelligence (Piaget 1963). These key elements of the learning process are understood to be strongly connected with human evolution and link cognition with its adaptive function. The processes involved in constructing knowledge develop in a cycle, so that an individual can constantly restructure and refine his/her understanding of the world (*organization*), and adjust to their environment (*adaptation*). When learners enter the state of *disequilibrium*, i.e. become aware of a disparity between their mental structures and what they have observed through experience (*assimilation*), those mental structures are modified (*accommodation*) so that the new observations could be accounted for and the state of *equilibrium* (i.e. agreement between observation and the learner's cognitive structures), is restored:

> The mind primarily assimilates, that is it perceives and categorizes experience in terms that are already known. Only if the result of this process causes a hitch and creates a perturbation, a review is initiated that may lead to an *accommodation*. This is to say, it may give rise to change in an existing structure or the formation of a new one. This (…) principle provides a mechanism for learning and should therefore be of interest to teachers of all kinds (von Glasersfeld 1997: 301).

The above cycle of revisions can be well applied to the concept of *interlanguage* (Selinker 1972), especially in the situation of advanced language learners, who gradually develop deeper and deeper understanding of the target language and are compelled to redefine the categories and rules they formed at earlier stages of their interlanguage de-

velopment. What is even more relevant to the topic of this study is that constructivism maintains a clearly defined, positive stance towards errors/mistakes. They are seen as a necessary and useful part of the process of adaptation that is inherent in learning:

> Mistakes inform the learning process enormously and enable a better understanding of the domain or concepts worked on – in other words, mistakes illuminate the learner and help him or her to learn and become more adapted to the experience or situation lived. (…) Mistakes are sources of learning and adaptation, and because of that, they should not be perceived negatively (Proulx 2006).

This does not mean, however, that errors should be ignored or encouraged. What is recommended here is an open-minded attitude on the part of both the learner and the teacher, and the readiness to employ mistakes in the process of learning, mainly by evoking the state of disequilibrium, when the learner becomes aware of a contradiction between received input and the cognitive structures s/he has developed. It must be added that experimental studies (e.g. Elliot and Devine 1994) have confirmed the effects of disequilibrium (also referred to as *cognitive dissonance*) to be extremely motivating for humans. The teacher's task is then to utilize learners' errors in the construction of new knowledge, and to build upon them in the most effective ways. This is exactly what the author of this study aims to do. From error analysis, through materials design to classroom activities, all the stages of the project strongly depend on the notion of disequilibrium and on allowing students to learn from their own, and their peers', mistakes.

Discovery learning activities recommended within the DDL approach are fully in agreement with the ideas of learning that Piaget's cognitive constructivism proposed. The learner assumes the role of a researcher and has access to raw data, which s/he needs to process to arrive at a feasible interpretation and build his/her own cognitive structure that would incorporate the newly obtained information. This is especially true in the case of hands-on concordancing, where learners have direct access to corpora and have a degree of independence in choosing the language problem they investigate. Still, even printed concordances, distributed in class for pre-planned activities, offer learners an opportunity to perform their own analyses of language data and observe regularities in them, which are common activities performed by language researchers. Such activities give learners a stronger feeling of control and autonomy, and involve their attention more effectively than traditional, transmission-based models of classroom teaching.

Table 6. Cognitive Constructivist Teaching Practices and Principles and their implementation in DDL (adapted from Bonk and Cunningham 1998: 33)

CONSTRUCTIVIST RECOMMENDATIONS	DDL PRACTICE
1. *Mind:* The mind is in the head; hence, the learning focus is on active cognitive reorganization.	Discovery learning requires learners to build new cognitive structures or reorganize those they have developed earlier.
2. *Raw materials:* Use raw or primary data sources, manipulatives, and interactive materials.	Concordances offer raw language data that learners need to process.
3. *Student autonomy:* Ask students for personal theories and understandings before any instruction. Allow student thinking time to drive lessons and alter instruction based on responses. Place thinking and learning responsibility in students' hands to foster ownership.	Learners obtain a new instrument in their language study (concordancer), which gives them more independence and allows them to assume the role of teachers' co-researchers rather than students.
4. *Meaningfulness and personal motivation:* Make learning a personally relevant and meaningful endeavor. Relate learning to practical ideas and personal experiences. Adapt content based on student responses to capitalize on personal interests and motivation.	DDL instruction tends to concentrate on those areas of L2 which pose particular difficulty to learners; this increases their engagement and gives them focus.
5. *Conceptual organization/cognitive framing:* Organize information around concepts, problems, questions, themes, and interrelationships, while framing activities using thinking-related terminology (e.g. classify, summarize, predict).	Concordance-based activities are naturally focused on concepts or grammatical relationships; they also force the learner to observe language, seek regularities, put various occurrences into logical groupings and use metalanguage to classify them.
6. *Prior knowledge and misconceptions:* Adapt the cognitive demands of instructional tasks to students' cognitive schemes, while building on prior knowledge. Design lessons to address students' previous misconceptions, for instance, by posing contradictions to original hypotheses and then inviting responses.	DDL is most suited to address problems already identified in learners' interlanguage and aims to eliminate or minimize them.
7. *Questioning:* Promote student inquiry and conjecture with open-ended questions. Also, encourage student question-asking behavior and peer questioning.	Questioning is especially encouraged when learners are given direct access to a concordancer and can formulate their own queries; printed concordances also stimulate learners to ask valid questions about form and meaning relationships.
8. *Individual exploration and generating connections:* Provide time for the selection of instructional materials and the discovery of information, ideas, and relationships. Also includes (sic!) encouraging students to generate knowledge connections, metaphors, personal insights, and build their own learning products.	A lot of corpus-based activities require sufficient time to be given for analysis of the data provided; the teacher needs to have an open mind and be ready to accept various observations, not only those he/she has made.
9. *Self-regulated learning:* Foster opportunity for reflection on skills used to manage and control one's learning. Help students understand and become self-aware of all aspects of one's learning, from planning to learning performance evaluation. Given the focus on individual mental activity, the importance of cooperative learning or peer interaction is in the modeling of and support for new metacognitive skill.	In most DDL activities students are encouraged to first analyze the material on their own and then share their observations in pairs or teams. Thus they learn to analyze language and observe others doing so. In this way they can develop awareness of the language learning process and improve their own performance.
10. *Assessment:* Focus of assessment is on individual cognitive development within predefined stages. Use authentic portfolio and performance-based measures with higher order thinking evaluation criteria or scoring rubrics.	Assessment of DDL activities can be performed on the basis of the validity of observations made by students, the complexity or validity of queries they formulated, and the regularity and intensity of their work with the corpus.

The table above places the teaching recommendations from cognitive constructivists next to some key characteristics of DDL and should make it easier to recognize all the convergent points. Other elements of constructivism that are prominent in DDL are learner autonomy and empowerment. With access to corpora and the development of skills necessary to use them, students become less dependent on their teacher in judgments of the appropriateness and grammaticality of L2 forms which they have used or consider using. Also, they can easily go outside the language range prescribed for their course and investigate questions of usage and grammar which for some reason have caught their attention or intrigued them. They have immediate access to language data and can become the teacher's partners in discovery rather than objects of the didactic process.

In the attempt to make students more autonomous in their language study it has become a common practice to encourage students to use dictionaries in class and check the meanings of new words which come up during the course of the lesson. However, studies analyzing the transferability of lexical knowledge acquired from definitions into active use or comprehension in new contexts and situations have proved this to be very limited (cf. Mezynski 1983, Cobb 1999b). Hence, students should be provided with more data, numerous examples of use of the target item in context, and frequency information. This is exactly what concordances can offer. Cobb (1999a) quotes several studies which prove the constructivist approach to vocabulary learning to be effective and the knowledge thus acquired to be transferable to new contexts:

> There are several mechanisms that could explain the link between multicontextual learning and transferable lexical knowledge. One is that the extra effort of drawing an inference makes the learning memorable (Hulstijn, 1992). Another is that meeting a word in several contexts paradoxically decontextualizes its mental representation, facilitating the computation of novel instantiations (Sternberg, 1987). Still another is that all cognitive and motor skills are schematized at an abstract and hence generalizable level if they are practiced in varying situations (Schmidt and Bjork, 1992) (Cobb 1999a: 18).

Although the studies quoted above do not test DDL techniques directly, they involve similar cognitive processes to those engaged in corpus-based activities and give ground for optimism about their effectiveness. The crucial evidence is the mechanism analyzed in Hulstijn (1992), where making inferences from context is evidenced to increase retention of the lexical items memorized. The downside of the process is also recognized in the study: students left to make their own generalizations without any

clues as to the meaning of the words tend to arrive at erroneous interpretations of their meanings. For this reason, it seems better not to leave the learner in control of the concordancer but to offer the materials in print; only then can the teacher provide sufficient commentary and guidance, and make sure that students do not become confused by the lesson.

It is worth noting that some of the findings of the third study cited above (Schmidt and Bjork 1992) may be seen as being in conflict with one particular aspect of concordance-based learning. According to the study, words are best remembered when revised in a random sequence of repetitions, intertwined with other items that are to be learned. Such sequencing proved to generate better retention than learning items *in bulk*, i.e. in longer sequences of repetitions of the same item. When this finding is considered in the context of DDL, it must be said that concordances present language items in the very fashion that Schmidt and Bjork (1992) consider less effective, i.e. *in bulk*. The key item is repeated in every concordance line and is practiced (or analyzed) in a continuous sequence. Doing it in any other way, however, (e.g. arranging lines from different keyword concordances in random order) is not feasible, as that would undermine the very notion of a concordance: the arrangement of examples in the way that makes regularities of usage clearly visible and distinct. Cobb's (1999a) point in quoting this study must be that the context in every concordance line is different, so it is not a mere sequence of repetitions but a series of interpretations of the same item in *different* contexts; hence the learning that results from the experience can be expected to be, after all, still effective. What is more, Schmidt and Bjork (1992) focus on practice and memorization, whereas DDL's use of concordances is most beneficial at the stage of presentation. Some forms of concordance-based practice do involve choices to be made between several lexical items, for example in a gapped concordance. These types of practice would be judged as the most effective following Schmidt and Bjork's (1992) criteria. It must be added that the 'standard' ELT techniques employed in the control groups in the study performed for this thesis also involved bulk presentation, i.e. one item at a time. This factor then could not have affected the outcome of the experiment.

What is most important in learning from concordances is the cognitive activity in which the learner is engaged while analyzing multiple examples of the target form. It would have taken a much longer time to encounter so much relevant input in the more natural, unplanned exposure to L2 written and spoken material, especially in the context

of formal *foreign* (rather than *second*) language education. Cognitive constructivism, originating from Piaget, and its concept of learning, seem to be in full agreement with how learning is seen by proponents of data-driven learning. The learner assumes the role of a researcher, and in DDL can actually use the same analytic tools as a corpus linguist, even if adapted to his/her needs. Constructivism, then, is the theory that is most often associated with these innovative, corpus-based language learning techniques.

2.5.2. Learning from examples – the neurophysiological view

Apart from cognitive constructivism, there are other possible views on the mental and neurological processes involved in data-driven language learning. The general concept worth mentioning here is the emergentist view of language acquisition (both first and second) where "knowledge is not seen as 'rules', nor is there any distinction drawn between 'declarative' and 'procedural' knowledge" (Ellis 2008: 465). The most influential model of this kind is connectionism, or neural networks theory, which views learning as a process of strengthening connections among neurons in the brain. This stands in opposition to the view of learning as consisting of rule formation based on serial algorithms, the type of processing performed by personal computers (Spitzer 1999). A neural network is a system of interconnected cells which send signals across the links between them (synapses) in order to perform an information processing task:

> We can hardly overemphasize the difference between information processing in neural networks and rule-based logical serial systems. The network, in contrast to the serial algorithm used by a conventional computer, contains neither rules nor calculation procedures. Its "knowledge" resides entirely in the weights of the connections. Although neural networks do not *contain* rules, what they can do can readily be described by rules. This distinction may sound sophistical, but it has far-reaching consequences for our understanding of ourselves (Spitzer 1999: 28).

The premise of the connectionist theory of language is that implicit language knowledge is stored in neural networks, and the rules of grammar only describe language and are not directly involved in its operation. According to Spitzer (1999: 29) "[s]uch rules as exist *are not in the head*, but are merely post-hoc ways of describing mental functions". Instead, connectionists believe language processing to be based on pattern recognition, a primary survival skill in the natural world, language being merely

one of its many implementations. Recognizing patterns is one of the brain's primary functions; it allows humans to group objects, phenomena and other entities into categories, to recognize their common features and be able to use them or react to them as they have learned to from previous experience. The same happens in the case of language, and is reflected in the concepts of *token* and *type* (see p. 84), for example, which are central to corpus linguistics. As Spitzer (1999) says, people learn from examples. It is not that each and every occurrence of a word, phrase or structure is memorized, but that they are categorized instead. The arrangement of the neural network is affected by a given instance only temporarily and then with time those changes fade. What is left is a general concept, or category, developed through our multiple exposure to it, the connections in the brain having been strengthened by those key aspects of the language item in question. This is the effect of a more general characteristic of how the brain works: wherever the world shows some regularities, the brain tries to recognize them, sometimes without conscious knowledge being involved. Unless there is some strong emotional involvement and personal reason to store a memory in the hippocampus, individual examples are not remembered, but are processed to build patterns on:

> In contrast with the hippocampus, the cortex is more like a "rule-extraction-machine". The synaptic connections between its neurons only change a tiny bit during a learning episode. That is why we are unable to remember most of our impressions later on. The fact that our brain does not record events as a video-recorder would, but rather extracts the rules underlying the events, is advantageous for several reasons. First, less storage capacity is needed if just the rules are stored and not every single event. Secondly, past coincidences are not useful for the guidance of future behaviour but past rules are (Spitzer 2005: 51).

The process described above is presented from the general neuropsychological perspective rather than being strictly connectionist, and hence the reference to *rules*. Connectionists would refer to processes of pattern recognition instead.

This is where DDL and its use of concordances can be invoked: the major advantage of a concordance as a way of presenting language material is that it makes recognizing patterns much easier. The arrangement of data around the key element in a pattern allows the learner to make observations that would be otherwise very difficult, if not impossible, to arrive at. Therefore, the format can facilitate pattern recognition and help build the connections in the brain that are later applied in language use. Since simple exposure to concordance data may not be enough for the learner to acquire a given

structure completely and use it appropriately, other activities involving the pattern might need to be introduced, so that the learner be engaged in the active use of the target structure. Concordances can, however, prime him/her for its further processing and so accelerate acquisition. There do not seem to be any experimental studies involving neural networks and using concordances in language teaching, but since *pattern recognition* is the key word in both areas, the connection is worth investigating. Larsen-Freeman (2003) suggests that the connectionist view of second language learning can be particularly useful in accounting for the acquisition of multiword phrases, or *chunks*, already discussed in connection with the idiom principle. The co-occurrence of language elements in strings can be seen as a result of them being readily accessible in the speaker's memory, due to high exposure in L2 input.

Generally, connectionism stands in opposition to symbolic models of language processing, such as generative grammar or cognitive grammar (Christiansen and Chater 2001). The type of processing it describes is not linear, as rule-based grammar is (if *x*, then *y*), but parallel. This means that the processes take place simultaneously across the neural network, rather than in a logical sequence of step-by-step processing. Various aspects of the context trigger a particular 'setup' in the network, generating an outcome that has been previously experienced and therefore strengthened in particular interneuronal connections. This is known as parallel distributed processing (PDP). Such an approach can account for the speed with which language processes occur, much exceeding the conscious analysis of choices that are made in language production and perception. Experimental studies on neural network models have shown that such learning is effective and does, to some extent, mirror the processes involved in learning simple patterns of language (e.g. past tense forms in Rumelhart and McClelland 1987). What is more, the connectionist model accounts for the very dynamic nature of interlanguage: it is never stable but evolves constantly, being readjusted and rewired in response to the new data which arrive with every supply of input.

It must be said, however, that connectionist ideas have been strongly criticized by many linguists for being able to account for first-order regularities in language only, but not for its infinite complexities at multiple levels (Pinker and Prince 1988). Larsen-Freeman has her own, though perhaps less forceful, reservations: "As attractive as connectionist models are, they clearly do not explain all human acquisitional experience. No computer can be programmed to reflect human agency or intentionality. Computers

are basically passive; they are not goal directed" (2003: 82). The connections/symbols debate has not really been resolved, and there actually have been attempts to reconcile the two seemingly contradictory points of view, though with the balance shifting towards symbolic manipulation in language processing (cf. Marcus 2009).

2.6. DDL and other trends in foreign language pedagogy

Constructivism, connectionism and prototype theory are relevant to data-driven learning by giving it some general theoretical foundations and accounting for the processes it involves. Apart from these, however, there are various theories of language teaching and language acquisition that need to be considered to fully appreciate these techniques, still novel in an average language classroom. These include the natural approach (the least obvious connection, admittedly), the lexical approach, and the general trend in language pedagogy recognized as *form-focused instruction* (FFI).

2.6.1. Krashen and the Natural Approach

As the longer discussion of Krashen's Natural Approach in section 1.3.6. (p. 44) postulates, one of its key notions is high exposure to suitable TL input, slightly higher in complexity than the learner's current level of language development. Another important notion in the Natural Approach worth mentioning here is the fact that the learner is not to be pressured into producing output in the TL before s/he feels ready for it. Both of these ideas can be linked with what data-driven learning recommends: learners are offered input that is relevant to their needs and that they seem ready to process, and their task is usually to analyze the data and draw conclusions, with production exercises being rather limited and fragmentary. The problem areas addressed in DDL activities are usually selected through analysis of learners' production and are supposed to help them make progress in those areas which proved difficult or particularly useful to them. This was the case with what may have been the first set of corpus-based teaching materials, i.e. the task sheets developed by Tim Johns and named by him as Kibbitzers (Johns 1997, also cf. Johns 2000). These materials were prepared by Johns for one-to-one ses-

sions with his college students who came for consultations on academic English to the English for International Students Unit at the University of Birmingham. The materials were made available online for other learners' benefit, and have served as model DDL materials ever since. Although Krashen and Terrell (1983) did not approve of *correcting* learners' errors, introducing tasks which address them need not be recognized as such an activity. Corpus-based instruction can be seen as an exemplar of the *i + 1* model: the learner has acquired a given form and tries to apply it in language production, but is not fully successful at it. Concordance-based activities can raise the bar to a quite precisely measured height and help the learner make progress.

Admittedly, Krashen's idea of language instruction precluded form-focused instruction of any type, either lexical or grammatical. His distinction between learning and acquisition rejected the idea that learners' progress could benefit in any way from explicit teaching/learning of issues related to particular language forms and their use. Instead, the L2 learner was believed to be able to acquire language from being exposed to an appropriate amount and level of meaningful input and gradually developing enough competence to be able to participate in target language situations. In this aspect, DDL techniques are very distant from the Natural Approach, but the very concept of learning from examples and from context is definitely common to both.

2.6.2. The Lexical Approach

A trend that was very significant in the development of DDL was Michael Lewis's (1993) Lexical Approach, an innovative view of language learning which emphasizes vocabulary learning and teaching. One of the key notions in the lexical approach is *chunking* (Nattinger and DeCarrico 1992), i.e. putting individual words together into meaningful and useful 'bits'. These sequences of words – collocations and lexical phrases – are believed to be stored in the memory as whole units, which makes them more accessible in speech production and allows speakers to achieve fluency (Pawley and Syder 1983). Lexical memory seems to involve some redundancy: words are stored both individually – as they can be recombined productively in various phrases and structures – and as parts of independent lexical units, where they constitute blocks that are not analyzed into elements every time they are used, but are used as prefabricated

chunks. This notion has already been discussed above (see p. 94), but since the lexical approach depends strongly upon it, emphasizing the social aspect of the nature of lexis, it needs to be brought back into focus. Lewis (1993) refers to the following passage:

> What makes an expression a lexical item, what makes it part of the speech community's common dictionary, is (…) that it is a social institution. This last characteristic is sometimes overlooked, but is basic to the distinction between lexicalized and non-lexicalized sequence. In saying that a lexical item is a *social institution* we mean that the expression is a conventional label for a conventional concept, a culturally standardized designation (term) for a socially recognized conceptual category (Pawley and Syder 1983: 209).

Such an understanding of lexical items constitutes an answer to the question of idiomaticity: not every sentence constructed in accordance with the rules of grammar that are part of native speaker competence is recognized by native speakers as *natural,* or *native-like*. There are many potential sentences/utterances which would be judged as unlikely, marked, or in some way peculiar. *Native-like selection*, the term introduced by Pawley and Syder (1983) for this phenomenon, is socially determined and hence needs to be considered in terms of frequency. The more frequent a sequence of words in popular use, the more natural and idiomatic it is perceived to be. This is where corpus tools become relevant: Frequency analysis and concordancing can give straightforward evidence of idiomaticity in a given string of words, and constitute a source of dependable data for both research and language instruction. For advanced language learners, being merely grammatically correct is not enough. In order to achieve near-native proficiency in the target language, they need to aspire to as near a natural use of both grammar and lexis as possible. Corpus-based language teaching materials can provide the type of input that advanced learners need – authentic and contextualized, but organized.

What is more, Lewis (1997: 53) argues that L2 learners need *negative* evidence to prevent errors, which naturally occur in the learning process, from fossilizing:

> Learners formulate hypotheses about language patterns by making (conscious and unconscious) generalisations on the basis of the input. Inevitably they make some over-generalisations, assuming combinations are possible which are not sanctioned by general use. (…) Although natural input provides alternatives, these may not be noticed and over-generalisations may become fossilised. Someone has to say what is not possible and this is clearly a task for the teacher.

At the same time, however, Lewis is wary of teacher correction, as it may prevent learners from taking risks and testing newly formed hypotheses. This is where corpus-

based techniques can become useful again: the errors that recur in learners' language can be chosen by the teacher as a focus of another lesson. There s/he can offer relevant negative evidence by presenting concordances that will help the learners verify the rules they have arrived at. A similar idea lies at the root of this very study, in which L1-based errors, most likely fossilized, are addressed with corpus-based activities.

A natural consequence of the enhanced emphasis on the role of lexical chunks in language is the blurring of the boundaries between grammar and lexis: "Language consists of grammaticalised lexis, not lexicalised grammar. The grammar/vocabulary dichotomy is invalid; much language consists of multi-word 'chunks'" (Lewis 1993: vi). The importance of grammar as an element of a foreign language syllabus is then significantly reduced, at least until learners reach a fairly advanced level of proficiency in their target language. Not that it should be completely eliminated, which is sometimes pulled as a charge against the proponents of the lexical approach. But there is a change in perspective: grammatical structures, sentence patterns, and other sentence frames are of secondary importance. Their role is to glue together the lexical items into cohesive and coherent sentences. Although arrived at from a different angle, this statement is very close to the concept of lexicogrammar adopted within corpus linguistics by Sinclair (1991) and his followers.

Another aspect of the lexical approach that can be linked with corpus-based teaching and learning is the concept of grammar as a receptive skill. Lewis (1993: 9) recommends developing exercises that "encourage the perception of difference of meaning". The reasoning behind this is that learners need to obtain a lot of input and develop good understanding of differences in meaning between various grammatical choices before they are made to produce any utterances of their own. "Classrooms need to be input-rich" (Lewis 1993: 27). This is where the lexical approach converges with data-driven learning again; corpus data provide very rich input, and DDL activities are focused on processing it, arriving at generalizations and making good choices rather than generating large amounts of spoken or written output. The key objective in DDL is consciousness raising, which is expected to help learners achieve a higher naturalness in their language production. Admittedly, Lewis's primary preference for rich input was listening and in the case of DDL it is reading, but the emphasis on input rather than output is an important defining feature for both the lexical approach and corpus-based teaching. This is where the two depart most strongly from the recommendations of

communicative language teaching, according to which language production and the enhancement of STT (student talking time) in the classroom is an unquestionable priority.

Further, the lexical approach connects with data-driven learning in its criticism of PPP, the standard language teaching classroom procedure often associated with behaviorism (e.g. Lewis 1992: 6). While DDL proposes the "identify – classify – generalize" process, the recommendation from the lexical approach is a different paradigm: "observe – hypothesize – experiment". At the initial stage of the learning cycle learners are to be presented with language input. In the lexical approach this will most often be texts (mostly spoken, but also written) and various de-contextualized lexicon-building and lexicon-organizing materials, selected on the basis of a functional syllabus. DDL will provide learners with concordances and other corpus-derived materials at this stage. Next, learners use the data provided in order to construct their own hypotheses about the item in question, and to try to make generalizations and observations. These may relate to meaning, collocation, colligation, phrase structure or syntactic patterns. In the final phase, if learners are ready, there is an opportunity for experimenting, i.e. verifying their hypotheses about the target language. Lewis (1993) has a strong preference for spoken production at this point, with writing, especially of longer passages, being postponed till as late as possible. Such a cycle seems worth considering in corpus-based activities as well. As has been already said, DDL has its own paradigm (*identify – classify – generalize*). Some of its elements can be recognized in the alternative process suggested by the lexical approach, especially in the first two stages: *observe* and *hypothesize*. The third stage, *experiment*, exceeds the scope of data-driven learning, although this does not mean that DDL is in some way contradictory to what the lexical approach proposes. One may assume that the stage of *generalizing* will ultimately be followed by production, once the teacher becomes convinced that the learners are ready for it. This, however, was not included in the DDL recommendations from Johns (1991) and can only be conjectured. What is of importance here is the rejection of the traditional PPP model and acceptance of a different style of the classroom procedure: with emphasis on individual thinking and autonomous processing. Learners are expected to become more aware of their language processing and take responsibility for organizing their knowledge.

Finally, the view of the type of context that language learners' attention should be directed at is similar for DDL and the lexical approach. In both, the context is seen more in terms of co-text than situation. For DDL the choice is determined by the very nature of the tools it uses. Corpus queries provide just this facility: the key word or phrase with some surrounding text. Some corpus interfaces, e.g. BYU BNC, provide tags for each line in a concordance saying what type of text it comes from, and this is very useful. But it can hardly be taken as a reconstruction of the situation the text originates from. Since the major point of corpus analysis is to establish what *linguistic* environment a given form tends to appear in, the situation itself is not highly relevant. As for the lexical approach, the choice of co-text over situational context as the primary form of contextualization is determined by its focus on collocation and colligation. "Contextualisation means noting the *situation* in which the word may occur, but most importantly noting *the co-text with which it can regularly occur*. If context is seen as situation + co-text, it is the latter – co-occurring language – which is more important for language learning" (Lewis 1993: 103; original emphasis). Co-text gives the learner the necessary background, sufficient to understand the function that a given phrase performs.

Summing up, there are several convergent points between data-driven learning and the lexical approach. These include focus on collocation and naturalness, the merger of grammar and lexis into lexicogrammar, stronger emphasis on receptive skills, criticism of the PPP model of classroom teaching, and the primacy of co-text over situation as contextualization. What is surprising here is that there is not much cross-reference between key publications of the two trends in language teaching. Lewis quotes Sinclair on very few occasions, and when he does talk of corpora, he seems very distrustful of them, as in the glossary entry here:

> *Corpus* A collection of texts assembled to form the material for a scientific study of language. (…) Different corpuses may be constructed for different purposes and produce unexpected differences. Many contemporary corpuses have been designed for lexicographic purposes; while they always give useful information, it is a highly questionable article of faith that the evidence they reveal is equally suited to devising language teaching materials (Lewis 1997: 215).

This is a clear vote of no confidence from the author of the lexical approach to proponents of DDL. There are no further arguments provided for this negative attitude, but it

indeed must be strong, as Lewis does not include a single corpus-based task among the many recommended types of exercises and activities which he lists in his "implementation book". Interestingly, data-driven learning publications hardly ever quote Lewis, and if they do, it is almost always limited to his stance on the importance of collocation. The lexical approach as such, as well as its recommendations for the language classroom, is usually skipped over, as if there was no connection between the two. Still, the lexical approach shares many elements with data-driven learning and it seems that the two trends in language teaching could each benefit from using each the other's ideas.

2.6.3. Form-Focused Instruction

Form Focused Instruction (FFI) is not a single school of thought in language teaching. It is more of an area of study whose goal is to identify effective language-teaching practices. It has a long tradition and encompasses all levels of language: phonology, lexis, grammar, and pragmatics. A lot of research within this area has been done to "test the rival theoretical claims of skill-building and input-processing models of L2 acquisition by examining the relative effectiveness of production-based and input-based grammar instruction" (Ellis 2001: 2). Naturally, this sort of research is of great interest to both linguists and language teachers; it is also highly relevant to the discussion of data-driven learning and its effectiveness.

In contradiction to Krashen's non-interface position and his claims about unconscious acquisition being the only way of developing a foreign language competence, Richard Schmidt (1990) formulated his *noticing hypothesis* (see p. 48). It says that acquisition will not happen unless the learner pays conscious attention to a given linguistic phenomenon in the input (form and its associated meanings). Only then can input be transformed into intake. These ideas were further developed and refined by Robinson (1995). He associated three mental processes related to attention (*detection*, *attention*, and *awareness*, as defined by Tomlin and Villa 1994) with Schmidt's noticing hypothesis, and considered noticing in relation to short-term and long-term memory: "Awareness is critical to noticing, and distinguishes it from simple detection. Noticing is a consequence of encoding in short-term memory, and is necessary for learning. What is noticed may be subsequently encoded in long-term episodic memory (memory for per-

sonal experiences)" (Robinson 1995: 298). Noticing itself, then, is not sufficient for acquisition to take place, but it is necessary. Stern (1992) classifies this approach to acquisition under *analytic teaching strategy*, in opposition to the *experiential strategy*.

The noticing hypothesis has strong support among language acquisition researchers, especially from researchers under the form-focused instruction banner, as opposed to meaning-focused instruction. It seems that after the many years' dominance of communicative language teaching and its meaning-focused instruction, the opposite trends are gaining ground. Language learners who finish communicative language courses tend to be quite fluent and confident, but have problems achieving the higher levels of proficiency and accuracy that some of them aspire to. The key issue seems to be fossilization and low levels of negative feedback the learners receive in their language classes. Hence the interest in techniques that increase accuracy is returning, but possibly not at the expense of the achievements of CLT, i.e. improved fluency.

Form-focused instruction is a term which embraces a variety of classroom situations and pedagogical options. These have been categorized as follows:

Table 7. Types of Form-Focused-Instruction (after Ellis 2001: 17)

Type of FFI	Primary Focus	Distribution
1. Focus-on-forms	Form	Intensive
2. Planned focus-on-form	Meaning	Intensive
3. Incidental focus-on-form	Meaning	Extensive

Though slightly confusing, the labels "focus-on-forms" and "focus-on-form" introduced by Long (1988) have become generally accepted and define the scope of interest within this area of study. The former, focus-on-forms, "implies that the teacher and students are aware that the primary purpose of the activity is to learn a preselected form and that learners are required to focus their attention on some specific form intensively in order to learn it" (Ellis 2001: 17). In the case of the latter, student's attention is directed to a language element as it appears in the context of a meaning-focused activity. Ellis (2001) has added a further distinction within this category, based on whether the form/forms in focus have been preselected beforehand or not, as researchers have tended to differ in applying the label to both planned and incidental attention to form. These are evidently two very different situations in terms of both the materials and activities offered in class and hence there are the three major categories presented in

Table 7. The last column in the table refers to the distribution of the target form in the language input and/or output of the activity.

Data-driven learning can easily be assigned to the first category. The teacher usually decides to use concordances or other corpus-generated materials in order to address an issue that has attracted his/her attention in the students' earlier performance. Therefore, the primary focus is, indeed, the form in question and how it should (and should not) be used by the learners. The concordance format offers probably the most concentrated distribution of the target form possible, allowing learners a large amount of input and giving them rich data to process.

According to Ellis (2001), focus-on-forms encompasses the following set of options:

- explicit vs. implicit teaching;
- inductive/deductive teaching;
- structured input vs. production practice;
- functional language practice.

DDL activities address language problems *explicitly*, without trying to hide from learners that they are dealing with grammatical, lexical or phonological issues in the lesson. Advanced learners, who are usually offered this kind of activities, are mature enough to be able, and willing, to improve their accuracy and make the effort required to do so. It is hardly an option to use a concordance without making it absolutely clear what form is being focused on, since it is either enhanced in the middle of the page by the KWIC format, or made prominent in some other way by the task itself (gap filling, partial translation, frequency list analysis, etc.). The form is intentionally visually enhanced and is supposed to attract the learner's attention. Learners consciously undertake the intellectual effort to make observations, generalizations and lexicogrammatical choices. DDL activities are, then, as explicit as can be about focusing on language form.

The next choice, inductive vs. deductive teaching, is again quite well established for corpus-based classroom activities. The very idea of DDL is based on discovery learning (see definition on p. 137). Learners are offered language data (either in print or through a text analysis software interface) and are encouraged to analyze them and arrive at feasible generalizations. Johns (2002: 108) formulated the following slogan for the sake of his own students: "Every student a Sherlock Holmes!". It renders the character of corpus-based tasks very effectively: there is a lot of analytic thinking involved on

the part of the students, and they are all expected to arrive at some solutions to the questions posed. That is why the slogan has become very popular with DDL researchers and practitioners.

The next dichotomy – structured input vs. production practice – is perhaps slightly less definitive. The dominant factor will be the former, corpus data being organized in a logical and reader-friendly way, but some elements of practice do appear in classes employing data-driven learning techniques as well. The initial stages of a DDL lesson involve processing structured input, mainly in the form of concordances or word-frequency lists, but other activities built around the concordance format can involve elements of production, though usually of the controlled kind. These may involve mainly gap filling, error correction, or translation. For DDL, production practice is then associated with making good language choices based on the surrounding text in the concordance, which, admittedly, does not offer much space for creativity and the type of production tasks that are used in more standard, communication-oriented lessons.

The last option listed by Ellis (2001) with focus-on-forms activities is functional language practice. In these activities learners are put in situations in which they have opportunities to practice target structures. They are classified as focus-on-*forms* instruction (rather than focus-on-*form*) because "despite the apparent concern for meaning, the primary focus remains on form rather than meaning, and learners are aware that the purpose is to master accurate use through repeated use of the target structure" (Ellis 2001: 20). This element, though quite possible as a follow-up activity, does not really belong within the domain of data-driven learning.

Ellis (2010) adds another useful dichotomy in describing explicit FFI: proactive vs. reactive instruction. Each can be either deductive or inductive, but while the former focuses on preventing error, the latter is concerned with addressing errors when they are committed. Table 8 shows the different types of classroom activities associated with proactive/reactive teaching on the one hand, and deductive/inductive teaching on the other.

Table 8. Types of explicit form-focused instruction (after Ellis 2010: 441)

	Deductive	Inductive
Proactive	Metalinguistic explanation	Consciousness-raising tasks Practice activities • production-based • comprehension-based
Reactive	Explicit correction Metalinguistic feedback	Repetition Corrective recasts

The place of DDL in this system is within the proactive inductive instruction. Although DDL activities would generally address issues already recognized as error-prone, they cannot be conceived of as interventions, but are preventive in nature. DDL activities are pre-planned and do not normally arise because of the teacher's immediate reaction to a student's erroneous utterance. They involve planning, analysis of learner language and preparation of materials. The point of a DDL task usually is that there is an element in the target language which is identified as a common source of error in a given population of students, and giving focus to it is expected to prevent further errors from occurring. Even if some students in the class have not encountered the problem yet, they may be warned of it in advance, and develop an awareness which will help them make the right choice when an opportunity arises. Activities of this kind are also very useful in preventing avoidance, where, being vaguely aware that there is a problem with a given form but not knowing where the problem really lies or how to solve it, learners do not have enough confidence to use that form. Further, the more careful speakers, who tend to avoid taking risks, may feel encouraged to start using the troublesome word or phrase once the problem has been addressed and their doubts dispelled.

Overall, then, DDL activities can be recognized within Ellis's (2001) classification as focus-on-*forms*. Before the differentiation between focus on form and focus on forms was introduced and generally recognized, however, authors talking about activities directed at accuracy practice in general used the term *focus on form*. This is the case with Celce-Murcia (1985 and 1991), who makes other significant distinctions concerning the importance of form-focused instruction (in broad terms) which depend on a range of learner and instructional variables in a given educational setting (see Table 9). It seems that the learner variables and instructional variables that appear in the column on the "more important" side of the table are the very qualities that are expected of advanced college students who major in English. The description also applies to the subjects in the study reported in this thesis. This is also the type of students who are the

most likely recipients of the DDL 'treatment'. Although there have been some attempts to introduce corpus materials at lower levels of proficiency, it is in academic contexts that DDL techniques are used most often, partly because primary and secondary school teachers are not very familiar with them, or not very confident about using them with their students. Also, materials like these are likely to prove too big a challenge for younger and less advanced learners.

Table 9. Factors which determine the importance of focus on form (adapted from Celce-Murcia 1985: 4)[20]

	Less important <——— Focus on Form ———> More important		
Learner variables			
1. Age	children	adolescents	adults
2. Proficiency level	beginning	intermediate	advanced
3. Educational	pre-literate; no formal education	semi-literate; some formal education	literate; well educated
4. Learning style	holistic	mixed	analytic
Instructional variables			
5. Skill	listening, reading	speaking	writing
6. Register	informal	consultative	formal
7. Need/Use	survival communication	vocational	professional

Since most of the characteristics tend to coincide with one another in the same type of learner, an image emerges of a learner who would benefit most from form-focused instruction in general, and from DDT techniques in particular: a person of mature age, with an advanced proficiency level, a good education, and formal language use, as well as a focus on writing and professional setting. Stern (1992) added learning styles to the table (point 4 here), but this is one category for which it is difficult to predict the dominant trend in a particular class of students. Well educated adults can have a preference for either a holistic or analytic type of processing, for example. Learning styles are very individual and do not seem to be tied in with age, education, level of formality, etc. It is important to point out, however, that this characteristic is a factor that may, nevertheless, affect the learners' attitudes to form-focused instruction and its effectiveness for individual learners.

[20] Point 4 in Table 9, referring to *learning style*, was not originally in Celce-Murcia's version of the table, but was added by Stern (1992).

Conclusions

Corpus linguistics made it possible to conceive of language in a completely different, more scientific way. Its instruments enhance features of language that were less discernible before, especially collocation, colligation and semantic prosody. This information and its availability affects the understanding of processes involved in first and second language acquisition. Corpus linguistics has also inspired an innovative language teaching methodology: data-driven learning, which has introduced corpus information directly into the language classroom.

Its constructivist foundations give DDL credibility in the educational environment of modern schools, and offer teachers confidence in implementing the aforementioned techniques in the classroom. Other theories of learning are also relevant: DDL can be accounted for, to some extent at least, by the neurophysiological views on learning, and especially by the connectionist ideas on pattern recognition. Additionally, some theoretical approaches to language acquisition such as the Natural Approach, Lexical Approach and ideas associated with form-focused instruction (e.g. noticing hypothesis) share certain notions and principles with data-driven learning. The instruments which DDL offers have not been fully appreciated by practicing teachers yet, and it is a task for both researchers and materials developers to popularize them for classroom use. As some papers reviewed above argue, DDL does not need to be confined to academic settings, but can also find its uses in schools, at different levels of foreign language instruction.

Chapter 3: The principles and practice of data-driven learning

Introduction

In 1987 John Sinclair of the University of Birmingham published the first dictionary of the English Language constructed on the basis of corpus material: *Collins COBUILD English language dictionary* (Sinclair 1987). Since then language corpus has become a standard tool in lexicography, and now virtually all dictionaries depend on corpus data to some extent. More and more reference materials for language learners, including grammar textbooks or vocabulary study books, are informed by corpora as well. Among the former there are publications by Sinclair (1990), Biber et al. (1999), or Carter and McCarthy (2006); the latter include the *English vocabulary in use* series from Cambridge University Press, with books by McCarthy and O'Dell (2002, 2005, and 2008), the *CorpusLab* books (e.g. Barlow and Burdine 2007), and many more. Another corpus-related enterprise in ELT materials development is the publication of course books which are inspired by and depend on corpus data. The first book series of this kind was the *Collins COBUILD English course* (Willis and Willis 1989), built on the basis of "the lexical syllabus". According to Römer (2008: 114), this textbook was groundbreaking: "With its focus on lexis and lexical patterns, the CCEC responds to some of the most central findings of corpus research, namely that language is highly patterned in that it consists to an immense degree of repeated word-combinations, and that lexis and grammar are inseparably linked". A much newer textbook series which is advertized as highly corpus-based is *Touchstone* (McCarthy et al. 2005). All these materials use corpus information indirectly, as the source of authentic examples for particular language

items or as an indicator of the word frequency, level of formality, register, and other important aspects of word use that are considered when lexical items are selected for these publications, i.e. at the stage of curriculum planning.

A more direct involvement of corpora in foreign language learning and teaching began with the concept of data-driven learning proposed by Tim Johns in 1986. He worked at the University of Birmingham, the leader of corpus linguistics in Britain at the time. It was here that John Sinclair initiated and created one of the oldest electronic corpora, the Bank of English, now part of the Cobuild Corpus. Other early corpora, which are in use to this day, are the Brown Corpus (the Standard Sample of Present-Day American English), published at Brown University in 1964 (Francis and Kucera 1979) and the LOB corpus (the Lancaster-Oslo/Bergen Corpus Of British English) named after the three universities involved in its creation in 1978 (cf. Johansson 1978).

The diagram below presents the different ways in which corpora can be applied in language teaching – directly or indirectly. This thesis focuses on issues and activities which could be located on the right side of the diagram, i.e. direct applications, from both the teacher's and the learner's perspectives.

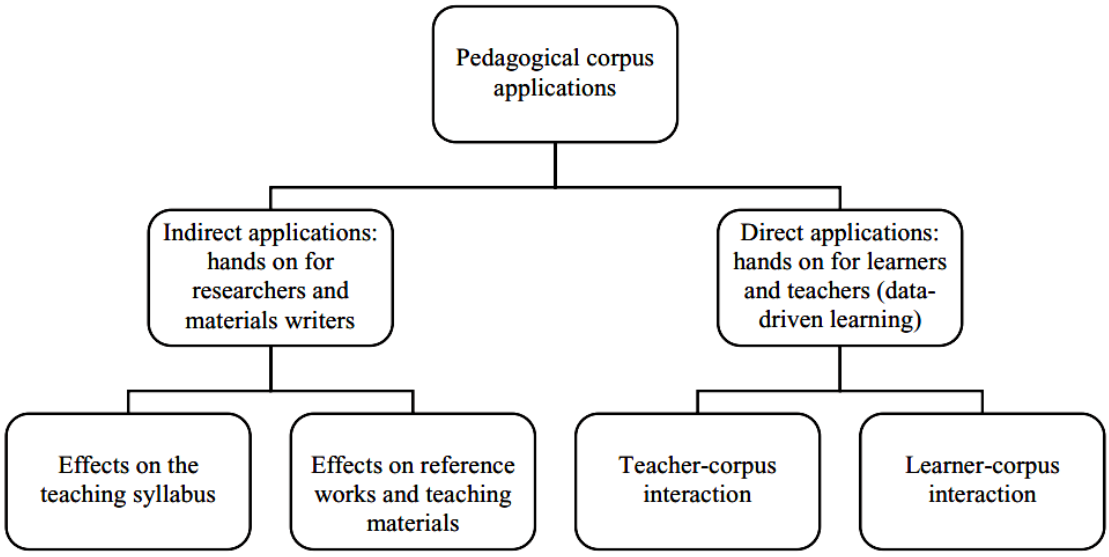


Figure 7. The use of corpora in second language learning and teaching (after Römer 2011: 207)

The aim of this chapter is to present the key concepts of data-driven learning (DDL), one of the newest trends in foreign language pedagogy, and the ways in which it operates. For this purpose, DDL techniques and materials are presented, some of which have been employed in the experimental lessons conducted for the needs of this study. Next, an overview of studies on the effectiveness of form-focused instruction and data-driven learning is provided, so as to provide some background information for the experiment carried out as for the needs of this dissertation. At the end of the chapter, the reader's attention is drawn to some shortcomings that the data-driven learning was found to suffer from.

3.1. Data-driven learning – definition and characteristics

The concept of data-driven learning arose as a natural consequence of the development and the general success of corpus linguistics. It was introduced by Johns (1991) as a particular kind of discovery learning based on corpus data. Johns explains:

> [T]he language learner is also, essentially, a research worker whose learning needs to be driven by access to linguistic data – hence the term "data driven learning" DDL to describe the approach. (…) The use of a concordancer can have a considerable influence on the process of language learning, stimulating enquiry and speculation on the part of the learner, and helping the learner also to develop the ability to see patterning in the target language and to form generalisations to account for that patterning (1991: 2).

Some variants of the term adopted for the methodology had been considered by Johns before "data-driven learning" was finally taken up by researchers and educators as the generally accepted label. Initially, Johns (1991) used the term *classroom concordancing* as parallel to data-driven learning, but then decided the former described merely the technique used in class, while the latter assigned the status of a methodology to his ideas (cf. Boulton 2011a). Each of the words in the term *data-driven learning* carries some important meaning and indicates a set of choices. The word *data* indicates the priority of authentic language over simplified, artificial examples generated just for the sake of teaching. *Driven* is understood to be stronger than *based* (Boulton 2011a), for example, and so again puts heavy emphasis on the role of real language facts obtained from corpora. The most important, however, is the choice of *learning* over *teaching* in the name

of the approach; in this way the inductive thinking of discovery learning is strongly accentuated, focus being given to the learner's cognition rather than the teacher's pedagogical procedures. Another label included in Johns' publications on the topic is *corpus-based language learning*, which appeared in his last paper (Johns et al. 2008: 495). The differences between all these are minor, and so the terms have been used interchangeably in this work. Some other complications connected with the use of the DDL label appear, though.

3.1.1. DDL and discovery learning

The definition of data-driven learning quoted above implies students' direct access to the concordancer and their own initiative in terms of what questions to ask and how to formulate their corpus queries. Such an approach allows the student greater autonomy and makes the learning process more rewarding. Classroom teaching, however, is not the best environment for this type of activity; apart from technical issues like access to computers, which is still far from standard in language lessons in schools, issues may arise concerning the degree of individual students' computer literacy, or readiness to use their own initiative in expanding their knowledge of and about the target language. For this reason, it is perhaps more advisable to introduce concordances in print first, and let the students become familiar with this format of language input without having to deal with any technical challenges. Once they have developed the skill of analyzing language in this way, they can be instructed on how to operate a concordancer, so that they could look for solutions to their own problems with English, dispel their doubts and learn how to use search syntax and the wide variety of options available. In this way corpus-based learning can enhance learner autonomy and the feeling of empowerment on the part of the student. In the classroom context, however, printed concordances seem to be a more manageable tool that can be more beneficial to a larger group of students. If the lessons are to be part of a research project, additional problems of normalizing the results will occur if participants operate concordancers on their own. The same approach is strongly advocated in Boulton (2008, 2010b and most of all 2010a), where it proved quite effective (see research review in section 3.4. below). What is more, according to Boulton (2009b), Johns was never orthodox about his definition, and used printed concordances

in his own teaching and research as well. In the current study, then, *data-driven learning* (DDL) is understood in a broad sense, as *any* language learning activity which is based on corpus data – not only concordances generated through learners' own corpus queries but also concordances, word lists and frequency data edited and printed by the teacher, provided to learners for the sake of consciousness raising. An extensive list of corpus-based activities and materials is presented and discussed in section 3.2. below.

There are many authors who support concordance-based activities as a useful way to implement discovery learning. Leech (1994), for example, says that analyzing a language problem by looking at examples from real texts leads to a much better understanding of how grammar operates in real communication than even a detailed presentation of the problem in a grammar book. Helping learners realize that native speakers' production is not always predictable and sometimes contradicts the 'official' rules of grammar may help them become more open-minded and more involved in analyzing language. The experience may help them see the human aspect of language, its unpredictability and flexibility. Also, working with a corpus can help them become more autonomous, as they will no longer depend solely on their teacher in matters of grammar. Some teachers may find it a challenge, but the change in role from "Teacher Omniscient" to "Teacher the Seeker" (Leech 1994: 20) might actually improve classroom dynamics. Students are more likely to be willing to cooperate with a teacher who sometimes admits s/he does not know all the answers and is willing to look for them together with the students.

Leech (1994) also observes that inductive learning is intrinsic in receptive skills, as the bottom-up information processing involved in listening and reading naturally entails making conscious or subconscious generalizations about language. Productive skills, on the other hand, are associated with deductive learning and top-down processing. According to Leech, when using the target language, whether in speech or in writing, learners construct utterances/sentences following the rules that they have learned. Thus the process here is reverse: from rule to language production. The question is how to bridge the gap between the two. The rules developed through induction (especially if based on authentic materials) are not always identical with what learners hear in the language classroom and read in the grammar textbook. It is hardly satisfying for the teacher to say that what they have heard or read is simply 'bad English'. Leech's answer is the prototype approach to grammar, discussed in section 2.4.1. above. Some

forms (or their uses) will be recognized as prototypical, others as less so. Some will be recommended as natural and native-like, while others will be seen as idiosyncratic or unique in their context.

Hadley (2002) conceives of DDL as a compromise between a *product* perspective on grammar and understanding grammar as a *process*, the two concepts proposed and developed extensively by Batstone (1994). The former can be identified with the structuralist approach, where grammar is perceived as a static set of isolated forms which are assigned certain meanings and are governed by idealized grammatical rules. The priority in product teaching is control. In this approach lexis is usually of secondary importance and is presented as less orderly and generally a source of complication to the neat grammatical structure of language. The latter concept, of grammar as a process, is understood as *procedural knowledge* allowing the learner to formulate a message which expresses his/her intended meaning. The emphasis here is on language use and successful communication, the notion which lies at the foundations of communicative language teaching and tasked-based learning. Since according to Batstone (1994) each of the approaches has its benefits and its shortcomings, neither should be used exclusively or in its extreme version. He proposes his own compromise approach, which he calls teaching grammar as a *skill*, in which students are given tasks focusing explicitly on language *form*, but at the same time generating meaningful language *use*. For Hadley (2002), however, it is DDL that can offer a possible way of finding a balance between the two extremes: students pay attention to form (product) by analyzing concordances and learn from them by performing inductive analysis (process), which results in consciousness raising. At the same time, by being exposed to real-life data, they avoid the danger of oversimplification and excessive idealization, which is the key problem in product teaching. In DDL, learners do not merely receive knowledge, but actively discover the regularities of the target language, and develop their 'feel' for it, which is a desired outcome in process learning. The teacher is a guide in this endeavor. Also, the use of concordances emphasizes the dynamic aspect of language, its flexibility: "DDL draws from process teaching in that it sees grammar as a flexible system of recurring and interrelated prototypes rather than a static set of rules" (Hadley 2002: 107).

Through analyzing concordances, learners can observe both the idiom principle and the open choice principle at work: the key word appears in strings of words (concordance lines) which reveal possible patterns, and how the 'slots' in those patterns can

be filled by a range of semantically related words. For example, Figure 6 demonstrates a transitive structure in which the verb *raise* is complemented with nouns that can be grouped semantically into the following categories:

- financial (*money*, *funds*, *cash*, *more*);
- intellectual (*consciousness*, *expectations*);
- discourse-related/argumentative (*subject*, *matter*, *questions*, *requisitions*);
- physical (*dust*, *head*, *voice*).

Although a dictionary entry may offer similar information, the cognitive processing involved in analyzing concordance data is believed to make the learning more effective (see the discussion of constructivism above). On the basis of these data the student can easily observe not only the grammatical pattern (here: the transitive structure) but also the variety of meanings that the key word can communicate. Those meanings need to be discerned from the words with which the key word collocates, or with which it is combined into phrases. It is believed that the intellectual effort invested into such an analysis makes the word (or pattern) more memorable and more easily accessible in further language production. Pawlak also recommends corpus-generated materials as "one plausible way of enhancing the potential of inductive grammar teaching" (2006: 272). He emphasizes the two major strengths of such materials: authenticity of the language input obtained through concordancing and the efficiency with which learners can recognize patterns of language. At the same time, Pawlak suggests that concordances should be used in the classroom only occasionally. In his opinion they might be too challenging even for very advanced learners, especially if the material is not filtered or edited by the teacher prior to classroom use.

Pattern identification involves various aspects of the data provided. Gabrielatos defines several constraints which must be taken into account for generalizations like this to be accurate, and these are:

1. The *medium*; that is, whether the sample contains only speech or only writing, or both.
2. The *context* of use, that is, "the physical, social and psychological background in which language is used" (Gabrielatos 1999: 15). The main contextual elements are the topic, the writer's or speaker's purpose, the type of text or interaction, the audience or participants and their relationship.
3. The *co-text*, that is, the surrounding text or linguistic neighbourhood of the feature, as words and structures seem to both attract, and interact with, one another.
4. The *representativeness* of the sample; in other words, the collection of texts needs to represent a microcosm of the language use of the population under investigation.

5. The *size* of the sample; as the example demonstrated, language patterns may be too large and complex for a small sample to reveal adequately (Gabrielatos 2005: 10).

In the case of paper-based classroom materials it is the author of the handout (usually the teacher) who needs to consider all these factors when making a selection of concordance lines to be included in the material. Whenever relevant, students should be provided with some background information, so that their thinking would be well grounded in facts rather than impressions.

Advanced students should learn to use corpora for one more reason: even advanced learners cannot fully depend on their intuitions about the target language. Unlike native speakers, who in most cases can make clear judgments concerning grammaticality, learners will have problems with such decisions, and if pressured, would have to classify more complex examples as *indeterminate* (cf. Schachter, Tyson and Diffley 1974). According to Pawley and Syder (1983), what causes most uncertainty of this kind is not syntax but lexis. Advanced learners can achieve high levels of fluency and grammaticality, but rules of lexical selection remain a challenge for longer. Their output, therefore, even if grammatical, contains expressions "judged to be 'unidiomatic', 'odd' or 'foreignisms'" (Pawley and Syder 1983: 193). Such problems are not easy to overcome in speech, but, when writing, advanced learners should be encouraged to consult a native speaker corpus to verify their lexical choices, especially in terms of collocation, colligation, semantic prosody and style, so as to make their target language writing as natural as possible. With time, thus acquired knowledge will surely improve their performance not only in writing but also, eventually, in speech.

The *Common European Framework of Reference for Languages* (Council of Europe 2001, henceforth *CEFR*) stresses the importance of developing such competences in learners as are necessary for them to become independent in the process of language learning. Those competences are put under the label of "ability to learn" (*savoir-apprendre*), and are described as follows:

> In its most general sense, *savoir-apprendre* is the ability to observe and participate in new experiences and to incorporate new knowledge into existing knowledge, modifying the latter where necessary. Language learning abilities are developed in the course of the experience of learning. They enable the learner to deal more effectively and independently with new language learning challenges, to see what options exist and to make better use of opportunities. Ability to learn has several components, such as language and communication awareness, general phonetic skills; study skills; and heuristic skills (*CEFR*: 106).

Such heuristic skills are particularly relevant to DDL: they include the ability to observe and draw conclusions from the analysis of authentic language material, the ability to find and process new information, and "the ability to use new technologies (e.g. by searching for information in databases, hypertexts, etc.)" (*CEFR*: 108). Paper based concordancing may be merely an introduction to developing those skills, but it certainly helps learners become aware of such possibilities and encourages them to become active and independent in their learning.

There is another advantage to the use of corpus data in language learning and teaching: being able to address issues in learners' interlanguage which result from L1 interference. As Granger (1998b) rightly observes, most EFL materials are produced for the international market and are seldom tuned to the needs of learners who are native speakers of particular languages. Instead, these materials focus on features of learner English which are common to learners of various ethnic backgrounds. Corpus-based materials can fill this gap. Kaszubski, for example, says the following about teaching English to Polish advanced learners:

> I am deeply convinced that assuming a Polish-specific perspective in the preparation of language materials will greatly benefit higher levels of EFL instruction, since many subtle nuances in L2 taught at this stage are better comprehended in relation to students' L1 and/or L1-based conceptual knowledge. (…) To ensure the optimum pace and efficiency, it is often best to customise the input intended for the learner – i.e. make it L1-oriented (1997: 137).

Such thinking lies at the core of the current research project: corpus-based analysis and corpus-based materials can be targeted at particular L1-dependent issues in participants' interlanguage, offering enhanced input which should help them to overcome those problems and to make progress in their English.

3.1.2. Key characteristics of DDL

Johns's (1991 and 1993) ideas established data-driven learning as a language teaching methodology, with the following distinctive characteristics:

> One of the most striking aspects of the methodology is the extent to which it mirrors the language-teaching Zeitgeist of the 1990s, combining as it does a central emphasis on authentic language use, on the development of discovery methods in the learner, and even (a goal much prized by our political masters in the United Kingdom) the possibility of restoring 'grammar' to a central position in language teaching (Johns 1993: 8, as quoted in Boulton 2009b: 6).

An important innovation lay in the fact that the methodology involved the computer in a new role: it became an *informant* rather than a *surrogate teacher*, as was the case with most CD-ROM language courses popular at the time. The computer, very much like a teacher, would pose a question to which the answer was already known and pre-defined. The user received feedback according to the typical classroom discourse pattern established in discourse analysis studies. Sinclair and Coulthard (1975) defined it as a succession of exchanges built on the following three turns: *Initiation* (the teacher asking a question), *Response* (students answering on the basis of what they have learned from the assigned material or classroom activity) and *Feedback* (the teacher responding to what students have said, evaluating their contributions against what the expected answer was supposed to be). This pattern is generally followed in the CD-ROM language courses and other computer-assisted language learning (CALL) materials, offering the learner a substitute for a real teacher and mimicking the classroom dynamic to some extent. There is usually a prompt or task, to which the user is supposed to respond, and then the computer generates feedback evaluating the response.

DDL is very different:

> What distinguishes the DDL approach is the attempt to cut out the middleman as far as possible and to give direct access to the data so that the learner can take part in building up his or her own profile of meaning and uses. The assumption that underlies this approach is that effective language learning is itself a form of linguistic research, and that the concordance printout offers a unique source for the stimulation of inductive learning strategies – in particular the strategies of perceiving similarities and differences and of hypothesis formation and testing (Johns 1994: 297).

The computer is used to generate a set of examples, either on the spot (if the student is given immediate access) or at the stage of preparation (if printed concordances are made available by the teacher). Then the data are to be examined in class, for students to arrive at generalizations on the basis of authentic uses of language included in the concordances. Such an approach makes learning more dynamic, and gives much less control to the teacher. The information learners receive is highly dependable: "Using this

approach, students can derive the information they need directly from the language, as though a computer were a tireless native-speaker informant, with rather greater potential knowledge of the language than the average native speaker" (Barnbrook 1996: 140).

Johns (1991) described his experience with concordance-based teaching in which students' generalizations and interpretations turned out to be more useful and convincing (as judged by the students) than what he, the teacher, had planned to say to account for the difference between *persuade somebody to do something* and *persuade somebody that* + clause. While his rule was based on purely structural analysis (the grammatical subject being identical for the second verb in the case of the infinitive, but not for the *that*-clause), the student observed the difference in meaning: infinitives appeared with verbs referring to actions, whereas *that*-clauses had verbs referring to truths and convictions. The students in the group all found their peer's explanation much more convincing and useful.

In cases like this, it is crucial for the teacher to be ready to assume an open-minded attitude and accept students' contributions. Only then will they enjoy the benefits of using corpus data in their learning. The corpus needs to be seen as an accessible tool, which can give learners the feeling of being autonomous in their learning, free to make their own judgments and observations. This does not mean that their conclusions are faultless: the teacher should be ready to point out those aspects of the data that have been overlooked, especially if they give evidence to the contrary of the learner's judgment. Also, it is the role of the teacher to guide learners in data analysis, especially if students find it difficult to see any regularities. The guidance can refer to such issues as whether they should look at the left or the right side of the node, what parts of speech they should be looking at in the data, or what contrast they should analyze. The teacher should, however, refrain from giving a ready-made answer to the problem. Instead, s/he can offer some metalanguage to re-formulate the ad-hoc rule that the students arrived at. In data-driven learning the computer does not substitute for the teacher in any way, then, but is a source of material that the students are to analyze, the teacher assuming the role of the "director and coordinator of student-initiated research" (Johns 1991: 3) rather than an omniscient supervisor. DDL techniques open opportunities for learners to pose genuine questions, the answers to which may prove surprising to both sides.

3.1.3. Identify – Classify – Generalize: the DDL classroom procedure

DDL recommends a classroom procedure that is alternative to the traditional PPP model (Presentation – Practice – Production). It is summarized in the sequence of a different set of three words: Identify – Classify – Generalize (Johns 1991: 4). Although Johns himself does not expand on the procedure, it is worth describing the process in some detail. The first stage, *identify*, requires the learner to observe what variations of form are being analyzed; for example, coming back to the problem quoted above, the choice between the infinitive and *that*-clause as a complementation of *persuade* is to be accounted for. The student needs to make note of such a variation in the data, and start looking for clues in the context as to what could trigger one choice or the other. Once they identify the trigger, they need to *classify* it, i.e. the distinction needs to be named, either in proper metalanguage or in the more informal terms which a student is more likely to use. In the case discussed above, the classification would involve distinguishing between verbs of action as opposed to verbs of truths and convictions. Finally, the learners need to *generalize* their findings; in other words a rule needs to be formulated which will bind the previously made observations together. The rule is usually formulated in conditional terms, stating that *if* a particular context appears (here: a verb of action), a given choice needs to be made (here: infinitive).

Johns (1991) points out one more effect that DDL can have on language learning: a new view of the role of grammar in language learning and teaching. Students' own queries may lead them to those areas of grammar which so far have been absent from pedagogical grammar materials or have been treated superficially in them. This might be the case for two reasons: some problems are avoided because of their complexity and high level of abstraction (e.g. the article system in English), others may be simply overlooked (e.g. transitivity). The latter problem seems especially suited to corpus analysis, which will easily reveal the pattern of nominal complementation in a transitive verb. Most of the grammar presented in reference materials used to be based on authors' 'armchair intuitions' rather than authentic data and were often inaccurate. Those intuitions concerned not only language use but also the selection of material for the course syllabus: what is and what is not problematic for the learner, what needs more focus and what can be treated more superficially or even left out altogether. Acceptability issues can be resolved by referring to representative corpora, and course

planning decisions can be supported with the use of learner corpora. Johns (1991) indicates the way in which grammar is introduced in language teaching materials as one of the reasons why its status was so low for many years, at the peak of CLT. According to him, grammar is often misrepresented in the teaching materials, in many cases making it look disconnected from how language actually operates. The use of corpus based materials "makes possible a new style of 'grammatical consciousness raising' (Rutherford 1987) by placing the learner's own discovery of grammar at the centre of language learning, and by making it possible for that discovery to be based on evidence from authentic language use" (Johns 1991: 4). Admittedly, things have changed since the publication of Johns's article (1991), and now more and more grammar materials depend on corpus information (for example Carter and McCarthy 2006 or Biber et al. 1999). Still, it is thanks to the ideas of data-driven learning that such publications actually appeared. What is more, even corpus-based grammars do not offer the same access to language as direct data analysis does. They may present an accurate and up-to-date analysis of the learner's target language, but the experience of studying raw language data is unique and engages the learner in intense cognitive activity, which even an excellent pedagogical grammar will not achieve. It is these mental processes that underlie the theoretical justification for data-driven learning.

3.2. Data-driven learning techniques – an overview

Over the years since DDL emerged as a new trend in language teaching, its proponents and practitioners have developed a battery of tasks and types of materials that can be employed in the classroom. Many of them have been described in academic and educational publications, although once teachers learn to appreciate the language corpus as a resource, they usually find the ways of using it with their students which are best suited to the language problem at hand and to their classes' needs. A commonly quoted catalogue of corpus-based classroom materials is presented by Tribble and Jones (1997), in what was probably the first purely instructional publication on DDL techniques addressed to language teachers. Many other publications have sections on concordance-based materials and activities, but most of them are of an academic nature: Johns (2002), O'Keefe and Farr (2003), Osborne (2004), or Lamy et al. (2012). Almost all

paper-based activities are built upon concordances, which are edited, grouped together, and exploited in various ways for educational purposes. Johns describes the editing process as follows:

> The most important principle that has to be borne in mind in carrying out this work is that the inevitable process of selection should not distort the evidence – that is to say, the concordance extracts chosen should represent as far as possible the full range of linguistic and communicative features of the raw data (1994: 298).

As Johns (1994) further explains, the real picture of language use can be distorted in the process in two different ways. The teacher may be approaching the subject with some preconceived idea of what *should* be in the material and makes the selection so as to meet those expectations. Alternatively, the teacher could be making the selection on the basis of defensible pedagogical criteria (clarity of the surrounding context, for example), but as a result obtain a set of concordances that misrepresent the real use of a given item. There can be three ways of dealing with the problem of distortion through selection: to make a purely random selection, to make the selection on the basis of knowledge acquired from an extensive analysis of the topic and ensuring that the distortion is as minimal as possible, and finally assuming a selection criterion that does misrepresent the real use, but acknowledging this fact to the students and making sure they understand how the data have been affected.

The very process of producing corpus-based materials was outlined by Adolphs (2006: 113), who recommends the following three steps, as part of a DDL materials design exercise for her readers:

Identify your audience: The materials need to be adapted to learners' needs in terms of level (beginners, intermediate or advanced), the type of language (general or special purpose), and the aspect of language they need to address (lexicogrammar, discourse patterns, or literature, for example).

Identify a teaching need: The language point that is to be addressed in the materials needs to be clearly defined (a new vocabulary item, idiom or collocation, a grammar or usage problem, a matter of style or genre, a recurring error, etc.); also, it needs to be established whether the teacher is to design and print the materials or to prepare a task for students to carry out with direct access to a corpus.

Choose a suitable corpus: After defining the above features, the teacher needs to decide on the best corpus to find data in; the choices may be between a spoken or written language corpus, general or specialist corpus, native-speaker or learner corpus, etc.

These choices help the teacher decide what type of task would be best in the teaching situation, on the basis of what materials it could be built, and how to select the data to meet the learners' needs. Corpus-based materials (and activities associated with them) could be organized into the following categories:

- L2 native-speaker corpus concordances;
- L1 corpus concordances;
- parallel corpus search results;
- learner corpus concordances;
- learner corpus and L2 concordances combined;
- non-concordance materials;
- materials for hands-on concordancing in class;
- corpus-based CALL applications.

3.2.1. L2 native-speaker corpus concordances

The native speaker corpus is the obvious and most common source of data in corpus-based teaching. All the sources mentioned above open their presentations with discussing this type of data presentation format. Below is a list of some variants of native speaker concordance-based materials used for classroom needs.

3.2.1.1. A straightforward KWIC-formatted concordance

KWIC enhances the key word in the center of each line in the concordance for the sake of easier analysis of the contexts in which they appear (see Figure 8). The task is usually to analyze the material and search for any regularities which may transpire from the data; in many cases the teacher will offer some guidance or questions to direct the learners' attention to particular features of the context (e.g. semantic prosody, collocations or set phrases, features of syntax or morphology).

re largely "lay" . The women, the **majority, were** unmarried; the men, married
the population aged 65 + the vast **majority are** able to cope with outdoor mob
on to such questions. Perhaps the **majority are** right to submit themselves to
its stake from 20 to 49 per cent. **Majority** foreign ownership **was** not permitt
to win and to win well. Labour's **majority is** still precarious in a sprawlin
-Provera can cause our women. The **majority** of doctors **are** simply not aware o
literature of the period that the **majority** of it **is** unambitious, anti-experi
ning centre was opened. The great **majority** of refugees **are** legally recognize
received were for new books. The **majority** of requests **were** for fiction (47%
o sit through something which the **majority** of them **are** not capable of doing?

Figure 8. A *BNC* concordance for *majority* with forms of *to be* up to 4 words to the right of the node (used for addressing problems with subject-verb agreement)

Tribble and Jones (1997) suggest the same format could be used for a language awareness activity in which learners are to recognize from context what part of speech a given word is (e.g. *like* – a verb or a preposition?), which homonym (*bank* – a financial institution or waterside), or which meaning of a polysemous word is used in each instance (e.g. *wood* – a kind of construction material obtained from trees or an area with many trees). Davis and Russel-Pinson (2002) propose a variant of the KWIC-based task in which the search term is not one word but a lexical root, and different words built from that root through compounding or derivation are placed in the node position in the concordance (e.g. *fish*, *fishing*, *fisherman*, or *Fisher*). The aim here is to make word-formation processes more accessible to learners and/or enhance differences in meaning between words which are morphologically related and are, therefore, often confused.

3.2.1.2. Comparison between two or more sets of concordances

Learners can observe differences in meaning and/or use between similar or frequently confused words, e.g. *speak*, *talk*, *say*, and *tell*. Various language problems may arise here: collocation, colligation, syntax, or differences in meaning or semantic prosody. Barlow (2002) suggests comparing concordances for different registers, genres or other varieties.

3.2.1.3. A gapped concordance (type I)

Students are given a set of concordance lines with a gap instead of the key word, which they are supposed to deduce from the contexts (usually done at a later stage in the lesson as controlled practice). Tribble and Jones (1997) list a variant of this task where the gap is replaced with a nonsense word, whose meaning is to be deduced from context lines in the concordance. The authors recommend that such printouts be examined carefully so as to make sure the context provides sufficient information for students to accomplish the task successfully.

Fill in each gap with the SAME word:

Problems remain because of such	_____	as the weights and balances found in
officers will have to consider such	_____	as is available to them.
the manufacture of the items from such	_____	as the tools that were used and
victim identifying the assailant. Such	_____	as saliva and seminal fluid tests,
South Africa's future. Yet such	_____	as there is suggests that his love

Figure 9. Gapped concordance (type I); answer: *evidence*

3.2.1.4. A gapped concordance (type II)

In a variant of the format presented above, the gap is not made for the key word, but for a word that is in some way dependent on it, usually to be chosen from a limited set of options (multiple choice question). Johns (1994: 308) offers an example, in which the gaps are made for articles (*the* or *Ø*) in front of the key word *industry*, a "notoriously tricky distinction in English". The same format of exercise is recommended for remedial instruction by Scheffler (2008) in his corpus-based analysis of advanced Polish learners' use of English.

3.2.1.5. Several sets of gapped concordances

Each set has a different key word, and these are all provided in a list (Johns 2002). The students' task is to match each of the words with one set of concordances. This is a useful task for practicing collocations, prepositions or making right choices of near synonyms which require different grammatical, semantic, or stylistic contexts.

3.2.1.6. Split-sentences matching

Johns (1994: 306) proposed a challenging task of matching halves of (extended) concordance lines split at the key word (in the original example it was the conjunction *that* in the result-clause structure *so… that*). For the task to be successful the context needs to be extended beyond the average length of a concordance line, so that learners could arrive at complete sentences in the outcome of the exercise. The task focuses very much on meanings expressed with the target structure; it also exposes learners to authentic examples of language use. According to Johns, students found the exercise very enjoyable, especially when trying to contextualize the sentences which turned out to be "false matches". The task's merit seems to lie in making learners appreciate the communicative function of an advanced syntactic structure, but it could be used with collocations or other lexicogrammatical problems just as successfully.

3.2.2. L1 corpus concordances – partial translation

Concordances in the learners' mother tongue feature much less frequently in DDL materials, but can be very useful in dealing with L1-related issues and errors in the target language. For the teacher, an L1 concordance may be a valuable diagnostic tool with which regularities of L1 can be identified, rather than assessed by recourse to the teacher's intuition. This is especially important if the teacher is not a native speaker of the learners' L1. More importantly, however, such a concordance can be used with learners in the classroom, though probably only more advanced groups should be considered.

L1 concordances are used for tasks involving partial translation: Learners are required to offer the most appropriate translation equivalent of the key word (and sometimes its immediate context). L1 concordances can be used to address issues of the misuse of L2 words or phrases, often resulting from L1 interference (e.g. deceptive cognates or underdifferentiation – see Chapter One). The translation task forces learners to reflect on their choice (and use) of L1/L2 equivalent words and it develops their awareness of differences between the two languages concerning the categorization of reality into concepts. For example, students often translate the Polish word *stwierdzić* as *to state* even though the context indicates arriving at a solution to a problem rather than making a statement (as in *naukowcy stwierdzili, że… – scientists have found that…*). Phonetic similarity between the two words is a likely source of confusion here. Another example can be seen in Figure 10, with a concordance of various inflectional forms of the Polish phrase *jakaś książka* ('a book'). The students need to decide which determiner to use in the translation: *a* or *some*.

How would you translate the phrase in the center in each line?

nie chciało mu się iść. Wziął	**jakąś książkę**	i czytał postanawiając sobie
wiedziała, co robić ze sobą. Wzięła	**jakąś książkę**	i siadła na werandzie, ale czytać
gablotach obiekty surrealistyczne,	**jakaś książka**	, jakiś dokument... A wszystko to
ciekawy program w telewizji. Może	**jakaś książka**	? A może po prostu - sen. To
że dzieci powinny mieć do nauki	**jakąś książkę**	. Jeżeli ja mam im rysować na
lalki, wszystko to przyda mi się,	**jakąś książkę**	przywieź też, o cokolwiek proszę
Teraz jeśli kseruje Pani	**jakąś książkę**	to korzysta na tym tylko posiadacz
Czy napiszesz jeszcze	**jakąś książkę**	? - Moim wielkim marzeniem jest

Figure 10. A task in partial translation: *jakiś* vs. *a* and *some*

3.2.3. Parallel corpus search results

Parallel corpora (i.e. multilingual aligned corpora) can offer engaging ways of introducing remedial instruction into the language classroom, especially in cases of L1 interference:

> Using parallel texts allows language learners to directly investigate (perhaps in response to queries posed by the teacher) the main correspondences between particular words and structures in two languages. The exposure of students to this information helps the formation of new schemas and the forging of appropriate schema-meaning links. (...) One problem that naive language learners have to overcome is the idea that there is a word-for-word equivalence between languages (Barlow 2000: 113).

Involving parallel corpora may be particularly recommended in training translators, but every advanced learner may need to approach the issue of translation equivalence, and should have an awareness of its complexity. As Partington (1998: 49) says, "complete equivalence" in translation is virtually impossible, and what should be aimed for instead is "adequacy of translation". In order to achieve this, learners need to be offered numerous opportunities to compare pairs of 'equivalent' words in L1 and L2 in context and make note of their referential meaning (to check for "cognitive equivalence"), their function in a given context ("pragmatic equivalence"), and their interaction with other elements of the language around them ("linguistic equivalence"). The most notorious in these respects are so-called *false friends* or *deceptive cognates*, defined by Granger and Swallow (1988: 108) as "pairs of words which are etymologically related, similar in form but semantically divergent". An excellent example of a task that addresses such a problem is St.John's (2001) comparison of the use of *also* in German with its equivalents in a parallel English corpus, *therefore* or *thus*. Partington (1998) suggests that working with parallel corpora can be more successful in preventing learners from making errors of this kind than depending on bilingual dictionaries, as the latter do not have enough space to include large amounts of contextual information for each lexical entry.

There are two major options with materials based on a parallel corpus: L1/L2 comparison and a combination of partial translation and comparison.

3.2.3.1. Comparison

Both the L1 and L2 results of a parallel corpus query are provided, and learners are asked to analyze the two versions of each passage to find and underline equivalents of the key word from the L1 version in the L2 text or the other way round. The task is intended to demonstrate to the learners the variety of available options in terms of equivalence and to let them become more open to the less obvious choices.

POL:	Nigdy nie zapomnę, jak pośród otaczającego koszmaru, bez **możliwości** zobaczenia jakiejś przyjaznej twarzy, przyszła moja kolej na oglądanie fotografii — nigdy przedtem nie czułem się tak szczęśliwy.
ENG:	I'll never forget it, sitting in the midst of a living nightmare **without** even a friendly face **in sight** and when it came to my turn to see the picture I looked at it and I never felt so happy in all my life.
POL:	Zauważyłem ją już wcześniej i, tak jak się spodziewałem, miałem dzięki niej upragnioną, choć mocno ograniczoną, **możliwość** obserwowania kilku metrów korytarza na zewnątrz.
ENG:	I'd noticed it earlier and, as I hoped, **it afforded me** a restricted but welcome view of a few yards of space on the outside corridor.
POL:	Nie spodziewałem się już żadnej rewizji celi, chociaż tutaj taka **możliwość** zawsze istniała i dlatego musiałem postępować bardzo ostrożnie.
ENG:	The chances of a cell search now were slender but **the danger** was always there so one had to be very careful.

Figure 11. Fragment of a parallel Polish-English corpus concordance for the Polish lemma *możliwość*

3.2.3.2. Partial translation and comparison

Initially, only the L1 version of the text is provided, and students are supposed to translate the key word and its immediate context into L2; then they receive the L2 text as it was retrieved from the parallel corpus and compare their translations with that text. The teacher needs to stress that they were not expected to obtain identical results, but draws their attention to the variety of ways in which the meaning of the original could be expressed.

3.2.4. Learner corpus concordances

A learner corpus offers a variety of opportunities for learners to address their problems with making the right lexical and grammatical choices. This is particularly true of trainee teachers, who need to be prepared to recognize, explain, and correct errors in the language they are going to teach. For the task to be most effective, the corpus should be built from texts written by students similar in their level, age, background, and other features critical to language learning. Some activities and materials based on learner corpora are listed and discussed below.

3.2.4.1. Error recognition/correction

A concordance shows examples of learner-produced language, the node being an item recognized earlier as one causing problems to the learners in class. Learners are supposed to decide in which of the lines the key word was used properly, and where it needs to be replaced or corrected in some way.

A comma in front of "that"*? Analyze the real examples of students' writing below and decide where the comma should be deleted or some other changes introduced. Ignore other errors, please.*

```
arguments. And it doesn't have to mean, that I totally agree or disagree with any
stic words? If you give up reading now, that will mean you are merely a brainless
ts flaws and tries to convey a message, that it's not the kind of faith that matte
ger than me this morning. The thing is, that in order to be good with children I n
 people, and gosh, I have internet now, that's really amazing ;) It's been very sa
of this. Therefore, a bypass is a must, that's a given. What I see as a gross dere
 of buttons and twinkling little lamps, that made a lot of noise and produced lots
thought to myself that I won't make it, that I can't stand it. My eyes were dry, b
 tequila shot and a burrito'. Wait, no, that's Mexican... I'm confused... Which on
of this stagnant state. The problem is, that he only does films about this one sin
I like to think of myself as a liberal, that's why I'm glad that democrat won the
d head full of ideas, but the thing is, that I do not believe in great changes. US
```

Figure 12. An error correction task based on a learner corpus concordance

3.2.4.2. Lexical enrichment

A lexical enrichment task can be used to develop learners' vocabulary and prevent them from overusing some words, for example for stylistic reasons. Osborne (2004) gives an example of the word *important*, which is the key word removed from a learner corpus concordance. Instead, the learners are required to choose a better word from a list of synonyms (*major, leading, wide, strong, severe, crucial, and established*). Thus learners are offered input with richer and more precise vocabulary so as to develop sensitivity concerning their choice of words.

3.2.5. Learner corpus and L2 native speaker concordances combined

Putting learner data next to target language native speaker concordances could reveal valuable information about learners' problems and, at the same time, offer a possible way of solving those problems.

3.2.5.1. Comparison

Learners receive two sets of data: from a native speaker corpus and from a learner corpus; their task is to compare the two sets and to identify systematic differences between them. These need not only be errors, but also cases of the overuse or underuse of a word, phrase or structure. In this activity learners can develop an awareness of how a given form is used by native speakers, and what uses should be avoided. The task was proposed by Osborne (2004: 261).

3.2.5.2. Native or non-native?

This is a recognition task in which learners are supposed to decide which set of concordances comes from a native-speaker corpus, and which from a learner corpus. "The objective is to develop critical linguistic distance, and to increase overall sensitivity to the characteristics of native and non-native writing" (Osborne 2004: 260).

3.2.6. Non-concordance materials

Apart from materials built on the basis of concordances, DDL instruction can employ other corpus data to produce ingenious and engaging language teaching materials, especially for advanced level learners. Most of them depend on various kinds of frequency information derived from a corpus.

3.2.6.1. Word frequency lists

Most text analysis computer applications can generate frequency lists for a whole corpus, its sections, or even a single text. Analyzing such lists with students can help them understand the differences between different styles characteristic of particular genres, for example, or differences between the English they use and native speaker English. Numerous 'official' frequency lists addressing different learners' needs have been published, but the one probably most popular in advanced classes is Coxhead's (2000) Academic Word List (AWL). The 3000-word list was built on the basis of a corpus of academic texts in English (3.5 m words) and gathers words that are common to various academic fields, but which at the same time lie outside the most common 2000 words in English. The list constitutes a good point of reference for learners of English for academic purposes (EAP), especially as far as setting the goals and choosing priorities in vocabulary learning are concerned. The *Longman Communication 3000* (*Longman dictionary of contemporary English* 2009, henceforth *LDOCE5*) is addressed to less advanced students, whose aim is to learn those words which occur in English most frequently (they constitute 86% of written and spoken texts) and so are most useful in every-day communication;

3.2.6.2. Lists of collocates and phrases generated through advanced queries

Some corpus interfaces allow users to define complex queries and generate lists of words or phrases defined by morphological, syntactic or semantic features of their components (cf. Davies 2004-, Davies 2008-). Lists of collocates and other words occurring within a specified span of the key word (left and/or right) can also be generated.

Adjectives ending with *-ate*

How do you pronounce these adjectives? Circle the ones which do not follow the general pattern.
Some of these words can also be verbs – find which ones. Place a tick in the box on the right.
How are they pronounced?

	ADJECTIVE	Verb?
1	PRIVATE	
2	APPROPRIATE	
3	SEPARATE	
4	IMMEDIATE	
5	CORPORATE	
6	ADEQUATE	
7	ACCURATE	
8	ULTIMATE	
9	DESPERATE	
10	INADEQUATE	

	ADJECTIVE	Verb?
11	DELICATE	
12	UNFORTUNATE	
13	LEGITIMATE	
14	MODERATE	
15	ELABORATE	
16	INTERMEDIATE	
17	DELIBERATE	
18	FORTUNATE	
19	INAPPROPRIATE	
20	AGGREGATE	

	ADJECTIVE	Verb?
21	INTIMATE	
22	PASSIONATE	
23	SUBORDINATE	
24	UP TO DATE	
25	INTRICATE	
26	IMMACULATE	
27	INACCURATE	
28	INNATE	
29	TEMPERATE	
30	APPROXIMATE	

(The words are ranked by frequency of occurrence.)

Figure 13. A task with a frequency list of adjectives defined by a morphological feature: suffix [*-ate*]

3.2.6.3. Frequency comparison list

A frequency list of collocates can be generated for two key words, so that learners would compare the contexts in which the two words occur, and make generalizations on their meaning and use. This is best applied to frequently confused words, e.g. *security* vs. *safety*, and words which are similar (but not identical) both in meaning and in form, e.g. *specially* vs. *especially*. The lists can be ranked in two alternative ways: by straight-forward frequency (which is based on the simple statistic of a collocate in the corpus and can sometimes yield random, uninteresting results with numerous articles, determiners and other high frequency function words), or by a Mutual Information score, a statistical index which reflects the frequency of the *co-occurrence* of the words in a corpus. Using this index prevents high frequency words from coming to the top of the list unless they do occur unusually frequently within the range defined in the query.

What is the difference in meaning and usage between SPECIALLY and ESPECIALLY?
Look at the words that most often appear on the right side of these. What is the pattern?

span: {0,1}

WORD 1 (W1): **SPECIALLY** (0.11)

	WORD	W1	W2	W1/W2	MI score
1	COMMISSIONED	55	1	55.0	496.9
2	CONSTRUCTED	46	1	46.0	415.6
3	DESIGNATED	16	0	32.0	289.1
4	CONVENED	13	0	26.0	234.9
5	COMPOSED	13	0	26.0	234.9
6	APPOINTED	12	0	24.0	216.8
7	TRAINED	60	3	20.0	180.7
8	ARRANGED	10	0	20.0	180.7
9	CHOSEN	19	1	19.0	171.7
10	DEVISED	9	0	18.0	162.6

WORD 2 (W2): **ESPECIALLY** (9.03)

	WORD	W2	W1	W2/W1	MI score
1	SINCE	245	0	490.0	54.2
2	AMONG	144	0	288.0	31.9
3	DURING	131	0	262.0	29.0
4	OF	193	1	193.0	21.4
5	THOSE	566	3	188.7	20.9
6	AS	669	4	167.3	18.5
7	IN	2278	14	162.7	18.0
8	AT	305	2	152.5	16.9
9	THE	1229	9	136.6	15.1
10	HIS	56	0	112.0	12.4

Figure 14. A task with a BYU-BNC list of words to the right of *specially* and *especially*. MI ranking.

3.2.6.4. Frequency bar charts

The BYU corpus interface can generate bar charts which represent the relative frequencies of a search term in different sections of the corpus. Such a graph can help learners grasp those aspects of word use which would be otherwise very difficult to demonstrate, and which are often beyond native speaker intuition. More importantly, learners can appreciate the flexibility of language and depart from the sharp right/wrong distinctions for the sake of the more fine tuned criteria of style, register and appropriacy.

*Look at these tables from the BNC. What do they tell you? Analyze the numbers for the use of "*fewer*" and "*less*" in front of plural nouns. When comparing popularity, look at PER MIL results (number of occurrences per one million words in the corpus).*
Is there anything in the data that surprises you?

query: less [nn2]					
SECTION:	SPOKEN	FICTION	NEWSPAPER	ACADEMIC	MISC
					
PER MIL:	6.0	0.5	1.1	2.7	2.3
SIZE (MW):	10.0	15.9	10.5	15.3	44.6
FREQ:	60	8	12	41	104

(academic section: *less wages, less resources, 100 or less,* etc.)

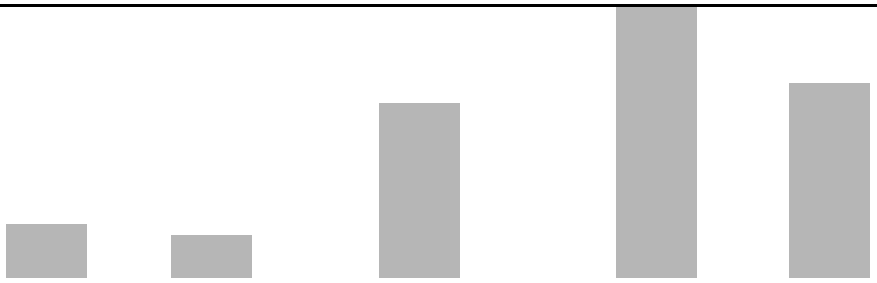
Query: fewer [nn2]					
SECTION	SPOKEN	FICTION	NEWSPAPER	ACADEMIC	MISC
					
PER MIL	4.0	3.2	13.3	20.4	14.8
SIZE (MW)	10.0	15.9	10.5	15.3	44.6
FREQ	40	51	139	313	658

Figure 15. A task with BYU-BNC frequency bar charts

3.2.7. Materials for hands-on concordancing in class

The focus of this work is mainly on paper-based DDL materials and activities, but brief information on the classroom use of corpora needs to be added. The key elements are the same: concordances, frequency lists, and other frequency data. Instead of data printouts, however, students are given instructions for operating corpora themselves and finding the information relevant to the topic of the class.

3.2.7.1. Worksheets with step-by-step instructions

A typical task is for students to investigate a given word, phrase or structure, and collect the data obtained from the corpus in a set of structured notes. Learners analyze the data and try to arrive at some generalizations. Feedback sessions based on students' individual or group work can reveal many valuable observations and findings, often surprising not only for the students but also for the teacher. An example of a worksheet like this is given in Figure 16.

Question: How acceptable is a construction like this:
The *reason* is *because* he didn't want to harm her?
Answer:
1. Open your WWW browser and go here: http://corpus.byu.edu/bnc/ . In the "Search string" field type "[pu*] The reason". In the CONTEXT slot write "because" – distance {0,7}. Press the "Search" button. Click on "BECAUSE" in the results window and wait a few seconds for the concordances to come up. Copy them and paste to an MS Word file. Then repeat the same for "that" in place of "because".
2. How many instances of "*reason* + *because*" did you find? _____
5. Look at the context of these instances by checking the corpus section label in the relevant line. What sort of text are these?

Figure 16. An example worksheet for hands-on concordancing
(adapted from Lamy et al. 2012)

3.2.7.2. Word usage profiles

Building a word usage profile is a challenging task, usually assigned to more advanced learners who also are quite proficient in operating a corpus and interpreting data. Learners are assigned a word and are supposed to study it in depth using the corpus. Then they need to produce a profile of the word, which should include information specified in the task. This could be:

- key collocates, left and right;
- phrases and idioms in which the word occurs;
- syntactic requirements and limitations;
- semantic requirements and limitations (co-occurrence with words carrying particular semantic features, such as ±liquid, ±animate, ±abstract);
- semantic prosody (favorable or unfavorable connotation);
- other unique features worth noting.

Needless to say, a project like this requires time and is worth considering as a homework assignment rather than class work. It gives learners an opportunity to develop their autonomy and contribute to one another's progress in a meaningful way. Therefore, a task like this is recommended with a very advanced, well motivated group of more academic-oriented students. Less independent learners would probably be discouraged by the challenge.

3.2.8. Corpus-based CALL applications

The overview of corpus-based materials closes with a brief discussion of computer-assisted language learning (CALL) applications. As aforementioned, most CALL materials are designed in such a way as to substitute the teacher in offering feedback to the learner, rather than to encourage the learner to become an independent explorer. Still, some of these programs do integrate corpus data into their tasks and engage learners in contextualized language analysis, which makes these applications more attractive and gives the user at least some degree of autonomy. Since such materials lie outside the main scope of this work, and are not very popular, only one example will be presented here in order to demonstrate the range of available options and mode of operation.

One of the earliest attempts at a corpus-based computer program for learning English was *PET•200*, designed by Cobb (1997) and named after the Preliminary English Test from Cambridge, for which Cobb's students needed to prepare as part of the English course that he taught at the time. The test had a lexical base of nearly 2,400 words which the candidates were supposed to know, and the program's major aim was to help them achieve this goal. The application included five types of tasks of increasing

difficulty, growing from less to more contextualized (from word-level to a text-level context), and from recognition to production. Here are some details of the project:

> All five activities access a 10,000-word corpus, which is simply 20 texts of about 500 words each assembled from the students' reading materials. The activities are driven by 12 wordlists of 20 words each, a total of 240 words over the term, or roughly 10% of the PET's 2387-word base. The 240 words were selected on the basis that they were unlikely to be known to the students, but likely to appear on a PET test, and occurred in the corpus at least four times (Cobb 1997: 304).

As can be seen in the description above, there is one DDL feature strongly involved here: the content of instruction is strictly matched with the learners' needs. The five following sections present tasks incorporated into *PET•200*.

3.2.8.1. Choosing a definition

For each of the 20 words in a session there is a list of 5-7 concordances, from which learners are supposed to deduce the meaning of the key word and choose one definition that matches their guess the best out of the four provided to them. For each concordance line there is an option of broadening the context, so as to provide more data for the learner to rely upon; there is also an audio file for each key word that the learner can play, to make sure the word becomes familiar not only in writing but also in its spoken form.

3.2.8.2. Finding words

Twenty words from the set defined for the session appear hidden in a string of random letters in a random sequence. Each item is accompanied by a set of gapped concordances, which are filled as soon as the learner selects with a mouse the correct sequence of letters to give his/her answer. The task helps students become familiar with the written form of the word as well as its use.

3.2.8.3. Spelling words

A sequence of 20 sets of gapped concordances appears, each accompanied by an audio recording of the word. The learner needs to type the answer for each set, and receives corrective feedback if the answer is only partially accurate. Students practice writing new words, but first need to consider the context in which they are used.

3.2.8.4. Choosing words for new texts

In the next stage students move from the word level to the text level. A few words – from the set they have been practicing in the session – have been removed from a text. Learners need to put them back in their original locations in the text. If they decide they need help with a particular gap, they can obtain concordances with the key word blanked out (the same as the one needed to fill the gap in the text). In this way they can see more contexts in which the word can appear and have more information on which to base their choice.

3.2.8.5. Writing words for new texts

The final stage is very similar to the previous one, except that now the words need to be typed rather than selected, and do not have to be given in a set sequence. The gaps can be filled with various forms of the target words, which makes the task more challenging and adds a grammatical aspect to the lexical focus of the task.

The five stages of the process move learners from the first exposure to the word, through recognition and focus on the form, to an active use of the correct form of the word in a meaningful, novel context. The transition is strongly supported by concordances at every stage of the session. *PET•200*, impressive as it was in its original form, was then developed into its more advanced version (*PET•2000,* Cobb 1999b). An important feature in the new version is a space for learners to add their own notes and word definitions or translations, which gives the program a more personalized touch.

Learners have more control over the sequence and number of words on which they want to work in one session, though the vocabulary load is assigned in larger sets, to be learned in a specified period of time.

The classification of DDL materials and tasks proposed above is one of many ways in which corpus-based classroom activities can be systematized. Krajka (2007b: 40) proposes a different method, based on what aspects of language are to be the focus of attention in a given task (e.g. "contrasting particular constructions, demonstrating their use in various contexts", "matching expressions with speakers coming from different geographical, social, professional backgrounds", or "verifying the prescriptive usage from the course book with the more contemporary, also geographically restricted, use"). The tasks are divided into four groups (grammar, vocabulary, reading comprehension, and writing) but the allocation to a particular group is often arbitrary and questionable. For example, is "assisting inferring new words from the text with additional, carefully selected examples" (Krajka 2007b: 40) an exercise in reading comprehension (as the author suggests), vocabulary development, or perhaps both? A list like this is certainly useful as a source of inspiration for teachers beginning to implement DDL in their lessons, but it is not very transparent. The point of working with a corpus is for the teacher to react to what happens in the classroom, to produce materials that address issues he/she can identify in a given group of students, and respond to the situation in a dynamic way, using a corpus as 'raw material'. Therefore, the classification of tasks according to its form and the type of corpus it uses seems more practical: the teacher will be able to see what instruments are at hand, what potential each of the sources has, and whether the data available there will be of use.

In the sections that follow the focus shifts from the presentation of data-driven learning and its instruments to the assessment of its pedagogical usefulness. Since DDL belongs to the general category of form-focused instruction, research on the effectiveness of this type of language teaching is briefly reviewed first, and then studies on DDL itself are discussed.

3.3. Research on the effectiveness of form-focused instruction

The effectiveness of foreign language teaching and its various approaches and techniques has been assessed in a wide variety of studies. The major controversy in the field centers around the effectiveness of form-focused instruction as compared with meaning-focused instruction. The research conducted around the issue is not always conclusive, especially in that quantitative studies will, by their nature, focus on form and accuracy; hence they will tend to assume the perspective and methodology of form-focused instruction.

For example, Loew (2001) analyzes the role of attention and awareness in language learning, addressing the validity of Schmidt's (1990) noticing hypothesis in an experimental study of learners' think-aloud protocols as well as their written TL production and some other post-exposure tasks. The results are reported as positive, indicating that "learners who demonstrated awareness of the targeted morphological forms during the experimental exposure appeared to have taken in and produced in writing significantly more of these forms compared with learners who demonstrated a lack of such awareness" (Loew 2001: 114). Norris and Ortega (2000) performed a meta-analysis of 77 experimental and quasi-experimental studies on the effectiveness of different types of L2 instruction in the years 1980-1998. It covers a wide range of instructional settings, but most involve English as the target language, at lower proficiency levels, with adult learners as participants in university settings. The project is described as follows:

> Within the research domain, L2 instruction has been operationalized as proceeding in terms of choices related to four components: presentation of rules, provision of negative feedback, exposure to relevant input, and opportunities of practice. Each of these four components presented multiple possible options for implementation and they could also be combined in various ways in a single instructional intervention, constituting particular pedagogical techniques (Norris and Ortega 2000: 462).

One of the five research questions posed by the authors concerned "the relative effectiveness of different types and categories of L2 instruction" (Norris and Ortega 2000: 428). This is a question that is particularly relevant to the topic of the current project, and so the results reported in response to this question are going to be summarized here. Unfortunately none of the studies included in the meta-analysis involved DDL

techniques, but at this stage the focus was on the overall effectiveness of the L2 instruction. Studies of the effectiveness of corpus-based techniques are going to be revised afterwards.

The outcome measure used for the meta-analysis of results in the study was Cohen's (1988) *d*-value, an indicator of effect size, which gauges the magnitude of differences between groups in standard deviation units. The advantage of this choice is that the *d*-value can be calculated on the basis of the most fundamental statistics provided by a majority of quantitative studies (group sample size, mean values and standard deviation values for the dependent variable in a given study). In some cases other calculations were necessary to arrive at the statistic, as *d*-value can be calculated from such statistical indicators as *t*-value (a *t*-test result) or *F* (an outcome of the analysis of variance – ANOVA). Overall, 45 out of the 77 studies included in the project provided sufficient data for calculating *d*-values for them. These 45 study reports provided data on 49 unique sample studies. The results relating to the effectiveness of types of L2 instruction are summarized in Table 10:

Table 10. Average effects and 95% confidence intervals for instructional treatment categories (adapted from Norris and Ortega 2000: 467)

Type of instruction	FonF (*k*=43)	FonF implicit (*k*=18)	FonF explicit (*k*=25)	FonFS (*k*=55)	FonFS implicit (*k*=11)	FonFS explicit (*k*=44)	All implicit (*k*=29)	All explicit (*k*=69)	Overall (*k*=98)[21]
CI upper (95%)	1.22	1.00	1.53	1.19	0.89	1.36	0.82	1.33	1.14
Mean	1.00	0.69	1.22	0.93	0.31	1.08	0.54	1.13	0.96
CI lower (95%)	0.78	0.38	0.91	0.67	-0.27	0.80	0.26	0.93	0.78

The meta-analysis indicates that there is no major difference in effect size for the focus-on-form (FonF) and focus-on-forms (FonFS) types of instruction. The major differences emerge between explicit and implicit attention to form. The former has twice as big an effect size as the latter, which gives a clear indication of its higher effectiveness and makes it a preferred choice for classroom practice. Ellis et al. (2006) have confirmed the higher effectiveness of explicit corrective feedback (metalinguistic explanation) for not only explicit but also implicit knowledge in an experimental study of the

[21] *k* indicates the number of instructional treatments, which is different from the number of unique sample studies (*n*=49) because one study can, and usually does, report on different treatments.

acquisition of regular past tense inflection. It is worth pointing out after Ellis (2010: 440) that "[t]he goal of explicit instruction is not just explicit knowledge but rather implicit knowledge, with explicit knowledge seen as just a starting point". This is known as the weak interface position, according to which explicit knowledge is not immediately transformed into implicit knowledge, but serves as a facilitator, helps learners focus their attention on language form and in this way supports successful acquisition.

The general hierarchy of effectiveness proposed by Norris and Ortega (2000: 465) on the basis of their analysis is the following:

FonF explicit > FonFS explicit > FonF implicit > FonFS implicit

The distinction between FonF and FonFS in their understanding of the terms, however, needs to be clearly specified, as it differs slightly from Ellis's (2001). An instructional treatment is classified as FonF (focus-on-form) if *any* of the first four strategies listed below (a-d) is employed in a study:

> (a) designing tasks to promote learner engagement with meaning prior to form;
> (b) seeking to attain and document task essentialness or naturalness of L2;
> (c) attempting to ensure that instruction was unobtrusive;
> (d) documenting learner mental processes ("noticing")
> In addition, many FonF studies presented evidence of
> (e) selecting target form(s) by analysis of learners' needs; or
> (f) considering interlanguage constraints when choosing the targets of instruction when interpreting the outcomes of instruction.
> (Norris and Ortega 2000: 438)

The category of FonFS (focus-on-forms) was assigned to studies which did not have any of the four key characteristics defined above, but directed participants' attention to a target form, usually designated through an arbitrarily assigned syllabus.

The types of activities that are within the range of DDL seem to be somewhere in between: the primary focus is on form, but at the same time learners try to establish how to use it in the most natural way and what meaning it expresses. There is always some context provided, though it must be said it is rather scarce and insufficient in terms of reconstructing the details of the situation in which the form was used. Definitely there is no focus on meaning in the sense that the learners themselves do not communicate their own message. What they do instead is try to reconstruct the meaning

intended by the original author of each concordance line. On the other hand, the 'extra' points (e) and (f) listed above fully apply to DDL techniques, especially in the case of the current study. Not only are they used as a result of a detailed needs analysis, which is standard practice for all DDL activities, but they also involve interlanguage problems, the special interest defined by the research questions posed here. For this reason, it is difficult to predict how this project would be classified by Norris and Ortega; it has some features of FonF and some of FonFS. In either case, it clearly involves an *explicit* focus on language forms and should therefore be found among the more effective types of classroom activities. Their unique characteristics connected with needs analysis and interlanguage discussed above give grounds to expectations of its having a higher effectiveness than other FonFS types of presentation and practice.

Studies on the effectiveness of form-focused instruction concentrate on various aspects of class activities. For example, Kim (2008) analyzed the connection between the intensity of students' effort invested in a vocabulary-learning task and its effectiveness. As could be expected, the relationship was found to be positive: those tasks that had a higher involvement load, which means they required more effort and/or more complex processing on the part of the learner, had a more durable effect. The study was based on Hulstijn and Laufer's (2001) involvement load hypothesis, whose experimental study also confirmed the positive relationship. If this finding were to be related to corpus-based activities, it could be said that learner involvement there is very high in terms of the mental processing, but rather limited in terms of the language production.

As has already been said, the way DDL techniques are used in language classes involves needs analysis and error analysis. The language problem to be analyzed is selected to address recurrent errors or other issues that jeopardize learners' success. Therefore, corpus-based materials often provide the language learner with negative evidence. The concept of negative evidence was developed in relation to issues of L1 acquisition and the question whether adults correct children's inaccurate language (cf. Brown and Hanlon 1970). Its commonly quoted definition is as follows: "Negative evidence occurs directly contingent on a child error, (syntactic or morphosyntactic), and is characterized by an immediate contrast between the child error and a correct alternative to the error, as supplied by the child's interlocutor" (Saxton 1997: 145). This notion has been used in the context of second language acquisition as well, and various studies have been conducted to assess the effectiveness of negative evidence as compared with

positive evidence. The general outcome of this body of research is to a certain degree inconclusive, with results varying for different language forms, different age groups and other learner-related factors (Strapp et al. 2011). The same authors conducted their own experimental study, however, whose conclusion is summarized as follows:

> As hypothesized, participants learned more irregular forms through negative evidence than positive evidence, replicating prior experimental research with L1 and L2 learners. (…) These results suggest that when negative and positive evidence are matched with respect to frequency and saliency, feedback that corrects a learner's error is more beneficial than correct exemplars modeled in the language (Strapp et al. 2011: 520).

It has been pointed out that the subject matter of the research is very difficult in terms of creating optimal experimental conditions for the results to be reliable. Strapp et al. (2011) attempted to achieve this by using artificially created words (nouns and verbs) and by training L2 adult learners on their irregular inflections by providing either positive or negative evidence. The design seems to be specially suited to the situation of advanced learners, who need to improve their accuracy by eliminating overgeneralization errors. As Strapp et al. (2011: 523) point out, it seems that advanced learners can benefit more from negative evidence, as it can "facilitate increased ability to use internalized forms"; these learners are more likely to have developed a necessary readiness for a given form to become part of their interlanguage. What is even more relevant to the current project, Strapp et al. (2011) report that negative evidence has been found useful in reducing transfer errors as well (cf. Tomasello and Herron 1989).

3.4. Research on the effectiveness of data-driven learning techniques

Data-driven learning techniques have been an object of numerous studies. Most of the experimental research of corpus-based activities has been collected and summarized twice by Boulton (2007 and 2010b). In both cases the body of research has been divided into three major categories:

- studies with the main focus on the learners' attitudes to the proposed techniques;
- studies of learners' behavior and practices when performing those activities (analysis of the teacher's observation sheets or of such self-report protocols as logs, diaries, interviews or class discussions);

- DDL efficiency studies, divided into two groups:
    - studies of corpus use as a reference tool;
    - studies focusing on the learning outcome of corpus-based instruction, i.e. its effectiveness.

The last category is strictly connected with the topic of the current study and it is this rather small collection of studies that is going to be discussed in some more detail here. First, a further distinction within this group needs to be made, however: there are studies that involve lessons with printed corpus materials and those with hands-on concordancing. The difference is not merely technical, as it affects the ways in which the lessons develop. Classes with direct access to corpora are more dynamic and focus on the process of obtaining information, while printed concordances allow more focus on the input and the language to be learned. Other differences have been discussed earlier (control of input, availability of equipment, planning of the lesson, etc.). Boulton (2011b), an online supplement to Boulton (2010b), provides an updated list of 93 experimental research projects on DDL, 29 of which investigate the learning outcome of DDL teaching procedures. Only nine (9) of these involve paper-based activities, the others having used learners' direct access either to a concordancer or to specially designed CALL software. The list of projects in Boulton (2011b) provides one more crucial detail for each of them: information on whether any statistical analysis is included or not, and in some cases it is noted that the quantitative data provided are very limited (raw scores and percentages only, for example). Out of the 29 learning outcome studies, seven failed to provide sufficient statistical information, out of which, in turn, two were devoted to printed corpus materials. As it transpires, the apparently large amount of research relevant to the topic of the effectiveness of DDL has been reduced to as few as seven papers, once the scope of research was limited to printed materials, and strict quality criteria were applied. These are:

- Sripicharn (2003)
- Tian (2005)
- Allan (2006)
- Koosha and Jafarpour (2006)
- Boulton (2008a) and Boulton (2010a)
- Johns et al. (2008)
- Boulton (2009)

The section that follows is a summary of each of those seven projects, with the focus on the results and conclusions relevant to the current study.

3.4.1. Paper-based DDL activities

The first included in the list is a study by Sripicharn (2003). Its design is very similar to the one used for the needs of this thesis, and involves two groups of students – experimental (*n*=22) and control (*n*=18). The former worked with corpus-based materials while the latter had classes which did not include any concordances or other corpus-generated input. The study's main instrument was a pre- and post-test, whose task was to measure the learning effect of the corpus-based elements of the observed lessons and to compare it with that of traditional lessons. The mean results of the post-test for the two groups were analyzed with the independent samples t-test, which failed to show significant differences between the two sets of results. Surprisingly, no attempt at calculating *progress* from pre-test to post-test was made, which puts the design of the study into question. Various possible explanations were offered as to why the experimental group did not achieve better results than the control group: pre-test results were high, indicating that there was not enough room for improvement (no specific data provided), the language input in the experiment was too varied (vocabulary, collocation, prosody, grammar and syntax), little time was allowed for the treatment (15 minutes for each session), or the experimental and the control treatment were not different enough. Finally, cultural differences were also thought to have had a negative effect on the perception of the inductive style of processing. Sripicharn (2003: 212) believes his Thai learners to be traditionally passive and accustomed to being "spoonfed by the teacher". Apart from the analysis of the learning outcome, the paper examines students' attitudes to the use of concordances in class by means of a questionnaire, as well as their performance in the concordance-based tasks. Both showed generally positive outcomes, with most students declaring that they found concordance-based activities useful and effective, though the group was divided exactly into halves over the question of whether they were interesting to do. In an open question posed as part of the study, the respondents were encouraged to make suggestions concerning the use of concordance-based activities. One of their suggestions was that these activities should be more strongly integrated with the

course they were taking and respond to their language learning needs, which seemed a problem in the reported experiment (items were generally too easy and not directly relevant to the syllabus of their writing class, within which the study was conducted). Another valid suggestion involved the role of the teacher: "The students feel they should be encouraged to find answers by themselves first, but in the end they still need confirmation and perhaps some conclusions from the teacher" (Sripicharn 2003: 222). All these suggestions were considered when the current project was being designed.

The next paper to be discussed is by Tian (2005). There is an experimental group (DDL) and control group (conventional teaching) in it again, and the pre- and post-tests are used to diagnose the learners' progress and to calculate the mean gain scores. The study analyzes the effectiveness of DDL in dealing with three types of language problems ("instructional focuses"): grammar, vocabulary and text-related features of language. In all these areas the gain scores were higher for the DDL group, but in the case of grammar the difference was not big enough to be statistically significant (as determined by an independent samples *t*-test). Apart from the major question, the paper addresses the issue of the relation between the proficiency level of the learner and the effectiveness of the DDL procedures, as well as the combination of the two (different language problems at different levels). For such a wide range of issues at stake, the design of the project itself is slightly disappointing in terms of sample size: the experiment designed to explore three broad areas of language is based on barely four language items – two grammatical ones, one lexical and one text-related. It could be said that a sample as small as this does not offer much ground for generalizations on such broad issues. The general design was similar to the one discussed above, except that students in both the experimental and the control group were further identified as either low- or high-proficiency students depending on their general English test results. This distinction was made arbitrarily on the basis of the pass mark being set at a 50% score in the test. Membership in either of the groups was sought to be related to the effectiveness of DDL activities for particular learners, but no such dependence was found in the results: "proficiency level made no significant difference in the students' gain scores" (Tian 2005: 366). The raw scores indicated a higher effectiveness of DDL in the "high proficiency" group for grammatical input and for textual features, and of conventional teaching in the lower proficiency group for grammatical and lexical input. The differences did not, however, prove statistically significant. In the final part of the paper the author

of the study makes an important point concerning the general benefits that DDL can offer to language learners, namely: contact with authentic materials and the development of inductive thinking skills. While the former may develop learners' general knowledge of the target language culture and thus help them learn some aspects of the language itself, the latter could have an even stronger effect, as the development of inductive reasoning skills would be beneficial in their further education, no matter what field of study they choose.

The next study that was listed by Boulton (2011b), and which fulfills the criteria defined earlier, is a paper by Allan (2006), which is slightly different to the two described above. First of all, the subjects of the experiment are diverse in terms of their L1s (and nationalities) and age (19-45). The context of the study was that of English as a *second* language, rather than a *foreign* language, which also distinguishes it from the research projects discussed above. The language-learning materials which constituted the key element of the experiment were assigned to students as part of their homework, which did not guarantee that the treatment actually occurred and that the conditions of the experiment were fulfilled. The learners were not college students as in the previous cases but evening course participants, who were preparing for their CAE exams. What the author calls the "concordance group" consisted of 13 learners, whereas the "control group" had only five (5) members. Furthermore, the treatment offered to the two groups can hardly be considered equivalent in terms of the amount and quality of input. While the concordance group was offered an extensive introduction to the aims and rationale of the project, the other was not offered such a session, which might have affected the learners' involvement, motivation, and, consequently, performance in the tasks. Additionally, the concordance group received regular 30-minute feedback on their performance in the concordance-based homework assignments over the 12 weeks of the experiment (every 2-3 weeks), whereas the other group was given two sessions of 15 minutes each, because apparently "learners asked relatively few questions during them" (Allan 2006: 24). This is hardly surprising, granted that no incentive to become engaged in the project had been offered to them: they were assigned conventional vocabulary tasks on the same lexical material, but were given much less background information about the project, and much less feedback and attention from its author. Considering all this and the minimal size of the control group, any comparisons made between the two

sets of results are bound to be problematic. Admittedly, the author of the paper did seem to be aware of these issues.

The general design of the experiment involved pre-test, treatment and post-test, as in the previously discussed study, the difference being that here the tests did not diagnose students' production as much as their assessment of how well they knew particular lexical items on the scale from 1 to 5 (the Vocabulary Knowledge Scale – VKS; Paribakht and Wesche 1997). The data in pre- and post-tests were analyzed with a paired *t*-test, to validate the effectiveness of vocabulary assignments in both groups. As could be expected, the experimental groups' average post-test score was much higher than the pre-test score, and the difference between them was statistically significant ($p < .05$). As for the control group, it must be stressed that its pre-test average score was much higher, actually higher than the *post*-test score for the concordance group. In such cases calculating gains/losses and comparing them is, by principle, questionable, as the conditions of the two groups' treatments are not well-matched. The overall gain score for the control group occurred to be negative, though its value did not prove to be statistically significant. In other words, the results indicated that during the experiment the five students' knowledge of the assigned vocabulary may have slightly decreased or remained unchanged; interestingly enough, it was still higher than the experimental group's average post-test score. The result may have been affected by the characteristics of the instrument applied: the losses appeared mostly in words which were previously given the highest rank (learners claimed their full knowledge, in terms of both grammar and semantics). The conventional materials which were offered to the students in the control group and which were hardly discussed with them may have undermined their confidence and caused a choice of a lower rank in the self-assessment test. This is the general problem of using such an instrument – the more advanced the students are, the more awareness they develop of the complexity of lexical knowledge, and hence the more cautious their self-assessment might be. A straightforward performance-based test would seem a more objective and dependable choice.

The study was supplemented with a questionnaire, in which learners in the concordance group could express their opinions on the DDL tasks, how much time they spent doing the tasks, and how interesting and useful they judged them to be. On the scale of 1 to 5, most learners chose level 3 in reference to their interest in the activities, and level 5 in reference to how useful they were. This seems to be a frequent outcome:

students see DDL activities as useful, but do not find them very exciting to do (cf. Sripicharn 2003, Boulton 2012, and the results of the present study). Some correlations were sought between the gain values and questionnaire responses about concordancing, but none were actually observed. The correlations that were found concerned (1) learner's interest in the activities and their opinion about their usefulness (r = .69) and (2) the amount of time they spent on the tasks and, again, their judgment of the tasks' usefulness (r = .49). These correlations reflect the consistency of the positive/negative attitude in the learners, but do not relate to the learning outcome of the experiment itself.

Generally, the study showed how careful one must be in designing experiments of this kind, and allowed the author of the current study to take steps that should prevent some problems from occurring. Much care was taken to offer both groups equal amounts of input and attention; the treatment was administered in class and monitored, and the overall conditions of the treatment for both groups were as similar as possible so that extraneous variables could be controlled or at least reduced to the minimum. The strong sides of the paper are its overview of research on L2 vocabulary learning and of data-driven learning methodology in its introductory part, and very detailed and thorough information on the experiment, its conditions and results. The author seemed fully aware of most of the study's shortcomings, and despite this decided to share all its details with her readers.

Koosha and Jafarpour (2006) published a paper on the use of print concordances in teaching prepositional collocations in English to advanced Iranian students. Its design meets very high standards: the sample is very big – compared to other studies in the field (60-item tests from 200 participants), the selection of students to join the experimental/control groups is random, the experimental and control treatments are extensive (15 two-hour sessions), and the statistical tests applied are fully suited to the type of data gathered in the experiment, which was not the case in some other studies discussed in this section. The problem here was that the first research question was formulated rather vaguely: "What is the role of the DDL approach in the development of collocational knowledge of prepositions among Iranian EFL students?" (Koosha and Jafarpour 2006: 197). The statistical analysis makes it clear that the authors were interested in comparing the effectiveness of DDL with that of conventional teaching, but the question itself is rather misleading. All the participants of the project shared their L1 (Iranian), and so the authors' focus (second research question) was also on how L1 col-

location patterns were carried over into their L2, in other words, on the effects of transfer. The third research question sought a correlation between the proficiency level and the degree of difficulty that prepositional collocations pose to learners.

The results of this project, especially regarding the first research question, are much different from the other studies discussed here so far, in that they are decidedly in favor of DDL. One-way ANOVA performed on the pre- and post-test data proved that the difference between the concordance group's and conventional group's results was significant, and that "the participants who took Data-driven instruction showed to be superior in the use of collocation of prepositions to those who received instruction in the conventional approach" (Koosha and Jafarpour 2006: 202). Unfortunately, no data apart from the ANOVA results are provided, which makes the outcome far from transparent and rather difficult to analyze. Neither are readers offered examples of the experimental treatment or the pre-/post-tests, so it would be difficult to replicate the study if anyone wanted to do so.

As for the question about transfer, the quantitative analysis of the material elicited from students through translation led the researchers to the conclusion that as many as 68.4 percent of errors in it were interlingual, and only the remaining 31.6 percent were intralingual. It must be observed that the study kept the two categories clearly separate, and did not consider the possibility of an error resulting from both L1 and L2 features at the same time. The 68.4 percent rate seems unusually high, compared with other studies of similar nature, which record from 3 to 50 percent of errors being attributed to L1 influence (cf. Ellis 2008: 355). In most cases the ratio approaches one third of errors, though the results may fluctuate depending on the tasks performed, the languages involved, or the learners' age and proficiency level. It is not entirely unfeasible, then, that prepositional collocations are in some ways uniquely L1-dependent. Finally, the analysis of differences in collocational knowledge between learners at different levels of proficiency was performed by means of the Scheffé test[22], and revealed that those differences do exist and are highly significant. Again, no details or raw data are provided, only the Scheffé test results, which makes the outcome of the study rather difficult to generalize. The authors then conclude that their results indicate the need to change the ways that prepositional collocations are taught so that there should be more emphasis on

[22] The Scheffé test is one of several post-hoc tests applied after ANOVA proves significant, to analyze differences between groups (cf. Hatch and Farhady 1982: 143ff.).

L1 as a point of reference and/or contrast. The third element of the study made Koosha and Jafarpour (2006) suggest that being strongly correlated with language proficiency, collocations should be an important factor in its evaluation. The paper, then, does make an important contribution to the development of DDL techniques and their use in the classroom, at the same time making valuable recommendations regarding teaching collocations.

The next study listed in Boulton's (2011b) review of research on the effectiveness of DDL study on less advanced students is presented in two papers by Boulton himself (2008a and 2010a). Both papers begin with strong argumentation in favor of paper-based concordance work in class, supported by statements about limiting the amount of novelty in one activity: "It is unsurprising that learners find it difficult to get to grips with new material (the corpora), new technology (the software), and a new approach (DDL) all at once." (Boulton 2010a: 539). The author also raises the problems of "technophobic" students who could be particularly uncomfortable or even frustrated with hands-on concordancing, and of unpredictable issues with equipment, website access, or some unexpected findings that can cause confusion. The design of the study involved three sets of five different grammar/usage problems identified as common sources of errors for 62 learners (in groups of 15-20) involved in the experiment. Each group was given a 60-minute experimental session in which one set of items was taught through DDL (printed concordances), one in a conventional way (dictionary-based materials mostly), and one served as control (the same in all the groups) and was not introduced at all. Half of the groups had items 1-5 presented to them through DDL techniques and items 6-10 conventionally. For the other half the treatment was the reverse. In this way all the students received identical types of treatment and at the same time could constitute a control for each other.

Each group's session was preceded by a 5-minute tutorial which made the notions of a corpus and a concordance clear to the participants. A pre-test was applied a week before the sessions, and three weeks afterwards participants were given a post-test. The test consisted of 30 questions, modeled on the multiple choice test of incomplete sentences in TOEIC reading exams (two for each item). Various statistical analyses were performed in the study, but those most relevant to the current project concerned the amount of progress between Test 1 and Test 2: a *t*-test (Boulton 2008a) and one-way ANOVA (with the post-hoc Tukey test) (Boulton 2010a). Since the latter is

more accurately reported on by the author, this will be the point of reference for the results of the study. DDL techniques proved slightly more effective than conventional methods, but the difference was hardly significant:

> A Tukey test derived from a one-way ANOVA conducted on the increase in scores between tests for each treatment (…) shows a minimal significant difference between the dictionary items and the corpus items (p=.15). In other words, although the DDL treatment was more effective than the traditional treatment, there is a 15% likelihood that this could be due to chance alone. The Tukey test shows a significant difference between the dictionary items and the control items (p=.013) and between the corpus items and control items (p= .0003).

In view of the results obtained in the current study, it is also interesting, and, admittedly, reassuring, that some negative results were obtained in that experiment, i.e. there were students who scored lower in the post-test than in the pre-test. This occurred for both DDL items (7 students) and conventionally taught items (12 students). It seems that drawing students' attention to common errors and problem areas, though generally beneficial and constructive, may in some cases cause confusion and affect the learner's performance in a negative way. It may be, however, that the effect is temporary, and the learner needs more time to benefit from the treatment. Awareness of potential problems may make it easier for learners to notice examples of appropriate use of the forms in question in real communication, and ultimately to acquire them successfully. For the sake of complete presentation, the remaining statistics on gains and losses were as follows: twelve students in each set made no progress between the two tests; the DDL results were higher for 43 participants, and, finally, conventional instruction improved the results of 38 students. The differences between the two sets are noticeable, then, but not particularly impressive, and as the statistics prove, not significant.

The participants[23] were also given a questionnaire to respond to, expressing their opinion on corpus-based activities and their usefulness. On the whole, the corpus work was found useful by more respondents than the dictionary-based work (59 vs. 31). More students declared that this type of activity can prevent them from committing certain errors (58 vs. 37). Also, more students expressed willingness to do more of such work in the future, with much lower numbers in favor of doing the traditional dictionary work (51 vs. 28). Questionnaire responses were tested for correlations with the students'

[23] The number of participants in the experiment is lower than the number of questionnaire respondents because of attendance issues.

TOEIC results and pre-test/post-test results. The outcome seems surprising: while quite a substantial correlation was observed between the general test and the post-test results for traditional instruction (r = .54), there is no clear correlation for the DDL instruction (r = .13). Boulton (2010a: 551) gives the following comment on this outcome: "[T]his might be interpreted as suggesting that all levels benefited equally from this type of information and approach, although clearly it gives a comparative advantage to the learners at lower levels of proficiency". More research is clearly needed on what type of learner responds best to DDL instruction, and this is one of the major conclusions of Boulton's two papers reviewed here. It is worth pointing out that both articles provide extensive information on the project in terms of relevant data as well as materials used in class, which is very helpful in interpreting the results and designing further studies on the topic.

The next project on the list is a complex experiment by Johns et al. (2008), where DDL was only part of a bigger scheme. The general purpose of the study was to test the usefulness of various CALL techniques in helping Chinese learners of English make progress in their language learning. Also, it was an opportunity to verify Krashen's (2005) theory of *free voluntary reading* (as referenced by Johns et al. 2008), according to which recreational reading[24] was the best way of developing reading comprehension skills, vocabulary and grammar. This appears to be an extension of the Natural Approach (Krashen and Terrell 1983), claiming that an appropriate type and amount of input is all learners need to make progress in their second language acquisition. Krashen (2005: 4) still believes focusing on language form to be pointless, and its results "peripheral and fragile".

The experiment described in the paper involved an experimental group and a control group, with 11 students in each. The former was given three hours of extra English (apart from the regular eight hours per week). In that time one hour was spent on reading assigned material and discussing it, one hour was devoted to various non-concordancing CALL activities and one hour was assigned to corpus-based work (L2 concordances in print, tasks with specially dedicated software, parallel concordances). The control group was supposed to use the same amount of time reading the assigned material (or other texts of a similar nature) at home. The differences between the effects

[24] Recreational reading is reading of extensive amounts of text relevant to students' needs and interests without any additional work being assigned on the material.

of the treatments were measured before and after the experiment with standardized English proficiency tests which the students took in their English course every term. The mean results of the test before the experiment were very similar for the two groups, except for the differences in standard deviation. The second test, however, showed some differences (78.91 and 76.82 respectively), which turned out to be statistically significant. The combination of intensive text work, CALL activities and DDL activities did improve learners' performance in comparison to the free voluntary reading programme. The corpus element was not assessed separately in the results, so the outcome of the experiment is not directly related to the current study. On the other hand, the fact that DDL element is so prominent in the design of that the research here confirms the growing popularity of the corpus-based approach and its perceived pedagogical value.

The latest of the articles from Boulton's (2011b) list which meets the criteria defined earlier (research on the learning outcome of DDL, with the use of printed materials rather than hands-on concordancing, and with statistical analysis rather than raw scores) is his own article on teaching linking adverbs to lower-level students by means of DDL techniques (Boulton 2009). Two major issues approached in this paper were, first, the dominant belief that DDL activities are only suitable for sophisticated academic students fairly advanced in their second/foreign language, and second, the need to have had some training in concordancing to be able to benefit from the DDL instruction (the author of the paper disagreed on both accounts). The design of the project was rather complex; it included four treatments (two corpus-based and two conventional), each with a different group, and three tests:

- pre-test;
- "reference-based" test with concordances available for consultation, performed immediately after the treatment;
- recall test after a ten-day delay.

Each of the tests had two types of tasks: ten sets of four gapped concordances and ten short contexts (between 1 and 3 sentences long) with one gap. In both cases the students were supposed to choose their answers from a limited set of options. The tests proved fairly difficult for the participants, so much so that nearly a third of the questions were left unanswered, but the former task seemed less challenging than the latter. The scores were significantly higher for gapped concordances than for short contexts (19.5% vs. 11.2% respectively, for the three tests performed).

The results of the study showed a relatively big decrease between the second and third test, which means that generally there is a large difference between the input and how much of it students are able to retain in their memory. Still, learners made progress as a result of the four experimental sessions, and the difference between test 1 and test 3 was significant for three treatments out of four (the exception was the one based on grammar/usage materials, which proved rather ineffective, especially in the reference-oriented test). The author concludes the study as follows:

> This study examined the ability of lower level learners to use authentic corpus data as a reference source and for learning. (…) Used as a reference source in this study, corpus samples led to more successful results than traditional pedagogical resources of the type the learners were familiar with: a bilingual dictionary and a grammar/usage manual. For the purposes of recall, the corpus and pedagogical resources were found equally effective. Of the two types of corpus data, it seems that authentic contexts in the form of multiple KWIC concordances are more amenable to lower levels than longer contexts consisting of one or more full sentences (Boulton 2009: 48).

The above overview of research on the effectiveness of teaching with printed corpus-based materials shows a complex picture, but some trends do emerge: in most studies DDL has some advantage over conventional activities, but usually it is not big enough to be statistically significant. Since corpus-based activities are often believed to be only suitable for more advanced learners, some of the authors tried to establish whether there is indeed a correlation between the learning outcome of DDL activities and the level of L2 proficiency. The studies reviewed here did not really confirm a clear connection between the two factors: as long as the language material in question is relevant to learners' needs and is presented in a student-friendly way, corpus-based activities may be useful and effective at all levels.

3.4.2. Program-based and hands-on concordancing – example studies

As stated earlier, the main interest of this work is the use of printed DDL materials. For the sake of comparison, however, reviews of one article on program-based concordancing and of another on hands-on concordancing have been included in this chapter as well. Thus it will be easier to appreciate the differences between the available

options, and the ways in which the other two ways of applying DDL, alternative to print-based concordancing, are implemented and researched.

The usefulness of concordancing with dedicated software programs in vocabulary acquisition was examined by Cobb (1999a). The tool used for the research was neither a concordancer nor a set of concordance prints, but a specially designed application called *PET•2000,* which offered learners a preselected set of concordances for each of the 2,387 words in the lexical base for the Preliminary English Test (PET), established by Hindmarsh (1980). The concordances originated from the 50,000-word corpus of course materials on which those lower-intermediate and upper-intermediate learners were working, so that the input they received would be suited to their level of proficiency in English. The control groups' practice was based on the same lexical material, but the students worked on bilingual dictionary definitions of the target vocabulary items rather than concordances. Both experimental and control groups had 12 weekly 45-minute practice sessions in a computer lab, during which they had access to their respective tools. The pre- and post-test consisted of two parts: Nation's (1990) "Vocabulary Levels Test" in which learners chose definitions for words randomly selected from the most common 2,000 words in English, and a gap-filling test, in which learners needed to apply the acquired words in new contexts. The results showed little difference in effectiveness of treatment on the first measure (choosing the right definition), but very impressive differences on the second measure (rational cloze), in favor of concordance-based practice. This indicates that the vocabulary acquired through DDL techniques is more readily available to learners in new contexts and renders these techniques more effective.

Finally, an example study based on hands-on concordancing needs to be presented; here students have direct access to a corpus interface and class activities involving their use of it. The example project chosen for this overview is Sun and Wang (2003). It is a small-scale experiment whose aim is to research the differences in effectiveness between inductive and deductive teaching of English collocations, the former with the use of online concordancers, the latter with conventional pedagogical materials. Two groups of about 40 participants each followed a typical pre-test/post-test procedure with four language items, two of which were easier, and two more difficult. The format of the tests was error correction. The choice of the format was supported by Woolard's (2000) statement that focusing on students' own mis-collocations could be useful in

raising their awareness of collocation in general. The same technique was used as a key element of the experimental and control treatments themselves.

Both the inductive and deductive groups were given short instruction on how their respective procedures were supposed to develop. The inductive group's initial step in the process was to identify an error in a sentence by searching for five examples of a given keyword through a web-based concordancer. Then they were supposed to observe regularities, identify a pattern for each of the items, and check the selected examples against it in order to confirm its strength. Finally, they were given sentences with errors to correct according to the rules they had established. The deductive group, on the other hand, was first presented with a rule and relevant examples, and then the learners were supposed to identify and correct errors in a set of sentences. At the final stage of the lesson both groups were given feedback on the proofreading task, so that they would know for certain what the errors really were.

The session was followed by a post-test immediately afterwards, and the results show a highly significant difference between the two groups in favor of the inductive one (p = .005). As far as the analysis of the results is concerned, however, the choice of the statistical instrument is questionable: despite having performed a pre-test, the authors of the study chose a one-way ANOVA and used it only on post-test results rather than the gain between pre- and post-tests. This instrument does not really account for progress achieved by means of the experimental and control treatments, but measures the differences between various groups in one dimension only (post-test results). The starting point of the experiment is not taken into account (or at least not reported), and there is no guarantee that the two groups were equal in the measured variable *before* the treatment. They may have started from different points, in which case the conclusions made on the basis of the experiment would be invalid. A different statistical instrument would be required for the research question concerning the effectiveness of DDL-style inductive teaching to be fully answered.

Still, the use of concordancing in class was well justified, and one can only wish the results of the experiment were more reliable as they seem unambiguous and clearly in favor of DDL techniques. Apart from the problems with statistical analysis, what may cause doubts is the concordancing activity itself. The students are reported to have had very minimal training in the use of the corpus instruments and so may not have been able to benefit fully from the experience. One could expect that they were given a rather

challenging task here, and with little support from the teacher or the task instructions. No feedback from the students is included in the study, so it is difficult to assess the experiment from their perspective. What is quite clear from the description of the procedure, however, is that there was not much control over how well students were able to operate the concordancer and how successful their searches were. The test results indicate full success, but, as has been pointed out above, their dependability is disputable.

The overview of research on the effectiveness of DDL techniques has been an opportunity to draw some conclusions concerning the design of the experiment for the current study. First of all, the research questions must be as precise as possible, because the limited number of participants available for the project does not warrant very broad conclusions. The choice of the measuring instrument (statistical test) must be made very carefully, to make sure the results are relevant to the questions posed, which was not the case in some of the research discussed above. Next, care needs to be taken to collect as much data as possible in order to avoid overgeneralizations. Although the number of participants is, in this case, limited by external circumstances, there are other ways of ensuring that there is enough information to process: the number of language problems analyzed, the number of test items per one language problem, and the number of sessions carried out as part of the experiment. These were limited too, however, because care had to be taken for the project not to disrupt the students' normal course curriculum, especially in view of their final examinations. Another challenge was to make sure that the control (conventional) classes were treated with as much attention and care as the DDL lessons, for the sake of objectivity of the study. Some of the aforementioned studies did not follow this policy, which undermined their credibility.

An important element of one of the projects summarized above (Koosha and Jafarpour 2006) was including learner errors in the design of the study, which was also crucial in the experiment carried out for the needs of the present thesis. There are several reasons why depending on errors was essential here: First of all, the students who participated in the project were training to become teachers of English, so the skill of recognizing and correcting errors was highly relevant to their needs. What is more, one of the tasks in their final exam in English was error correction, so they were highly motivated to improve in that area. For these reasons the task was included in the experimental and control lessons wherever it was considered useful and relevant. The exercise was expected to be even more effective and engaging if it could be based on errors

committed by the students themselves and their colleagues; such an opportunity arose with the idea that the written material produced by the students in their blog project could be accumulated into a corpus and incorporated into the experimental lessons.

3.4.3. The effects of learning styles

Learning styles have already been mentioned as an important factor in determining the amount of focus on form that particular teaching contexts require. Stern (1992) included the holistic vs. analytic dichotomy as one that should be considered in these choices (with the holistic style requiring less emphasis on FFI, and the analytic style – more). This is clearly not the only pair of options associated with learning styles. There are many others: the inductive vs. deductive style, field-dependent vs. field-independent learners, active vs. reflective, sensing vs. intuitive, visual vs. verbal, and sequential vs. global learning style, to name but a few. The last four pairs of categories are included in the *Index of Learning Styles Questionnaire* (Soloman and Felder 2012), an online tool for diagnosing learning styles for research purposes. Boulton (2009c) used this questionnaire in his study on how learning styles contribute to the variation in the effectiveness of DDL instruction. In it he refers to Kaszubski's (2008: 174) (as cited in Boulton 2009c) observation that learners' attitudes to corpus-based activities can be divided into three categories: "adopters, minimal users, and refusers", probably depending on their preferred learning styles. Then Boulton reports on research focused on particular learning style dimensions. The first one discussed is the deductive vs. inductive style, which appears to be most strongly connected with DDL instruction. Three studies (Lee and Liou 2003; Chan and Liou 2005, and Lewis 2005, as cited in Boulton 2009c) report that the inductive style gives learners a strong advantage in such activities. This is hardly surprising, considering that the involvement of inductive processing is the key characteristic of DDL. It is worth pointing out, however, that the majority of participants of the first two studies were found to prefer the deductive style, which the authors of the paper attributed to their cultural background (Taiwan). Still, talking in more general terms, Felder and Silverman (1988: 677) claim that "[i]nduction is the natural human *learning* style" while "deduction is the natural human *teaching* style" [emphasis mine, AL]. In their natural environment, i.e. outside the classroom, people learn from the

situations and problems they encounter, and they need to process the information they receive in order to arrive at some generalizations. Data-driven learning operates on similar principles.

Another dimension of learning styles that has been found relevant to corpus-based activities is field independence vs. field dependence. The former is associated with deductive, rule-based learning, and students with these preferences benefit more from the instruction which occurs usually in formal educational contexts (cf. Droździał-Szelest 1997). Students with the latter learning style, on the other hand, are more successful in inductive processing, prefer learning in interpersonal situations and are usually good communicators. The DDL techniques are not easy to locate on either side of the field-dependent vs. field-independent divide: they belong with formal, classroom teaching and so should be suitable for field-independent learners. Their inductive character, however, makes them a good choice for filed-dependent learners. The results of research on which of the two learning styles is more facilitative to data-driven learning are ambiguous: while Flowerdew (2008, as cited in Boulton 2009c) found in her study that field-dependent learners benefited much more from concordance-based lessons, Turnbull and Burston (1998, as cited in Boulton 2009c) came to the opposite conclusion. Their project, however, was a case study with only two subjects, so generalizations should be made with special caution. Generally, field-independent learners are thought to be more successful in learning foreign languages because their learning style facilitates "analyzing, restructuring, hypothesis testing and inferencing" (Droździał-Szelest 1997: 55), which are crucial abilities in language learning. It may be that DDL is unique in that it favors field-dependent learners, which would actually be a positive feature: such learners usually find formal instruction more challenging than field-independent learners do. It would therefore be extremely useful to have a set of classroom activities that meets their needs. The question whether DDL is more successful with field-dependent or field-independent learners needs to be researched more thoroughly for the answer to be more definitive.

Boulton (2009c) based his learning-styles study on the following four dimensions: active vs. reflective, sensing vs. intuitive, visual vs. verbal, and the sequential vs. global learning style. As mentioned above, these four were included by Soloman and Felder (2012) in their online questionnaire. One learning style proved particularly relevant to DDL, with statistically significant results: *visual* learners showed more positive

attitudes to this type of instruction than *verbal* learners. The correlation is quite powerful ($r = 0.42$), which is rather surprising, considering that the object is language, an evidently verbal entity. On the other hand, the visual enhancement of data in the concordance format may indeed have a strong appeal to visual learners, and help them make up for weaknesses in their verbal abilities. As far as the effects of instruction are concerned, the other three dimensions, active (rather than reflective) and sequential (rather than global) learners performed better, but the differences were not statistically significant.

This short overview of research on the effect of learning styles on DDL suggests that those students whose learning styles are in some ways unusual in their educational context benefit more from concordance-based techniques. Three such groups were identified: visual and field-dependent learners, both types believed to be less predisposed to foreign language learning than their opposites, and inductive learners in the Asian environment, by some researchers viewed as strongly attached to the *deductive* mode of instruction. These results can be seen as promising for those learners who do not respond very well to traditional, rule-based teaching.

3.5. Limitations of DDL

Like all teaching techniques, the use of concordances and other corpus data in the language classroom is not without problems. Most of them have been already mentioned, but it may be useful to gather them all together and assess their gravity. Krajka (2007a) lists the following issues:

- the lexical material in concordances may be unknown and so cause confusion;
- incomplete contexts in concordance lines may be incoherent and cause comprehension problems;
- using corpora efficiently requires some training and general computer literacy, without which learners may be limited to obtaining less informative simple word search results;
- due to the above-mentioned limitations, less experienced students may need continuous assistance from the teacher before they learn how to formulate corpus queries properly;

- some learners may not feel very well disposed toward inductive discovery learning, which may prejudice them against DLL activities;
- language material in corpus data is often unsuitable for classroom use as input, because it may be idiosyncratic and is not accompanied by any information about the speaker's status;
- the dominance of written language in corpora may be misleading to teachers in their judgments based on corpus information and lead to misinformation;
- the variety of formats, tagging systems, tools and interfaces in corpora available online may be a challenge for users.

The list of these charges is long and could be rather daunting. Nevertheless, most of these problems can be eliminated, or at least minimized, by using teacher-generated, printed materials rather than allowing students to access corpora directly. Incoherent samples in concordance lines, or those which contain unfamiliar, low frequency words, can be eliminated by the teacher, or adapted to students' needs without significant loss of authenticity. Technical issues, if they do arise, can then be faced by the teacher at the stage of preparation rather than 'live' in class, with many students calling for assistance. This allows the teacher time to solve problems without pressure and prevents possible confusion during the lesson. As for the idiosyncratic quality of the language material available in corpora, the answer is awareness: the teacher needs to be conscious of the differences between various language uses and to choose a corpus or sub-corpus that represents the variety of language most suited to the topic of a given lesson or learners' needs in general. It may sometimes be necessary for the teacher to supply some cultural background information for a given concordance line, or to reconstruct a possible situational context for it together with the students. This can actually add a valuable cultural aspect to the lesson. Recipients of DDL instruction are most often advanced learners, and it is crucial that they should be exposed to examples of language as it is used by real people, for all practical purposes. What is more, use of other authentic resources, either spoken or written, may cause similar problems.

There are problems which are not included in Krajka's (2007) list, but do cause some concern from the didactic point of view. The first to be mentioned is the low involvement of communication and creativity in DDL activities. Part of a typical corpus-based task is to formulate a rule on the basis of the data provided, which is done through negotiation in teams; the task, however, does not usually involve much interaction as

students do not engage in such conversations very actively. Learners who are eager speakers and prefer interaction to form-focused activities may indeed feel somewhat discouraged by DDL tasks. The results of the survey carried out as part of the study seem to confirm these observations. In cases like this, it is the task of the teacher to engage the reluctant learner as much as possible, and make sure that lessons include a variety of different activities, so that students with different preferences could find at least some of them attractive and engaging. This may lead to another general reservation from teachers: preparing a DDL-based class may seem time consuming and work-intensive. One can only say that most good lessons, especially those focusing on language forms and their use, need a large amount of preparation, no matter what technique they engage.

Some teachers may find working with corpora uncomfortable because it forces them to alter their role in the classroom: rather than the ultimate authority and the only judge of what is and is not 'correct', they become facilitators, guides and co-researchers. An independent, corpus-empowered learner may find information that undermines his/her teaching, which may be difficult for some teachers to accept and address. However, with such wide access to information as young people have these days, corpora are not the only source of reference that they will use. The mere fact of trying to look for answers on their own must be seen as a sign of high motivation and involvement on the part of the students rather than as a threat. This is the direction that modern education has been taking for some time, which makes DDL a very appropriate addition to a modern language classroom. Developing a sense of common learning and mutual benefit in the group will help solve such issues.

New solutions in language teaching often draw opposition from more traditional practitioners and researchers. In the case of data-driven learning, some of that opposition is justified, and requires to be addressed in classroom practice; most problems can be solved by preparing corpus-based materials in advance and making sure that there is a balance between different types of activities in the course – meaning-focused and form-focused, involving the whole range of language skills and addressing all learning styles.

Conclusions

The intention of the above overview was to present data-driven learning as a dynamic area of research and classroom practice, a field which has not yet reached its full potential and is open to improvement and new ideas. DDL instruments and activities are relatively new among various solutions for the language classroom. They are not meant to dominate in a language course, but can definitely constitute a valuable supplement to other, more meaning-oriented activities. Their value lies in the ability to reconcile two contradictory options in developing linguistic knowledge: *breadth* and *depth*. The former emphasizes quantity, and can be associated with learning vocabulary from word lists; it allows rapid growth in the number of words learners are familiar with, but their knowledge is usually very superficial, short-lived, and not easily accessible in normal communication (Cobb 1999b). The latter, on the other hand, stresses quality, and usually takes place when the learner encounters new language in context, through reading, listening or interacting with others. The problem here is that it is definitely less intensive, and more difficult to incorporate effectively into an advanced language course and to fulfill learners' needs. An additional challenge in the educational context is evaluation: testing learners on a list of words is much easier than testing them on lexicogrammatical elements which have featured in assigned reading material in unique contexts, in one of many meanings.

DDL is a compromise between these two choices. The teacher can plan a set of vocabulary or other language items identified as relevant to students' needs, perhaps drawn from the reading texts planned for the course, and then work on those items with the support of corpus-based materials, either in class or as part of homework. Perhaps the list will not be as long as topic-related vocabulary lists in advanced courses usually are, but the outcome should be much more satisfactory: the target forms will become familiar in their multiple functions and meanings (from different concordance contexts), and learners' knowledge will be deeper than that developed through memorizing lists of words. Learners will know key collocations, colligations and the semantic prosody of these items, and will be able to "place them meaningfully within various networks in relation to other words" (O'Keefe et al. 2007: 54). All this should not only make the new material more available for active use but also help learners become more independent in their language development.

Despite the benefits discussed above, corpus-based instruction has not yet seen its full application in language classrooms, especially in schools. Few teachers are even aware of the technology and its potential; many are not computer literate enough to be able to use it. It is therefore important to demonstrate DDL techniques to young teacher-trainees, who perhaps one day will decide to use them in their own teaching, with their own students. It may seem that a considerable amount of research has been done in the area, but it hardly affects teachers' daily work. A quite recent statement from one of DDL experts seems to confirm the impression:

> I would, however, still be hesitant to say that corpora and corpus tools have been fully implemented in pedagogical contexts and would argue that much work still remains to be done in bridging the gap between research and practice. The practice of English language teaching (ELT) to date, at least, seems to be only marginally affected by the advances of corpus research, and comparatively few teachers and learners know about the availability of useful resources and get their hands on corpus computers or concordances themselves (…). In addition, current language-teaching materials still differ considerably from actual language use as captured in corpora (Römer 2011: 206).

One of the conclusions that could be drawn from the review of the papers on the effectiveness of DDL presented in this chapter is that it is crucial for a study to be transparent. The reader should receive full information about the conditions of the treatments, the materials and processes they involve, as well as the results obtained. Otherwise it would be impossible to duplicate the study and verify its outcome. Also, such information makes it easier for the reader to understand how the research was carried out and to interpret its results. The next chapter provides detailed information about the research project undertaken for the needs of this thesis, including the procedures, instruments and materials involved, as well as the results of the study.

Chapter 4: The effectiveness of DDL in the advanced English classroom – a study

Introduction

In order to verify the assumed usefulness of using corpora in teaching English at an advanced level, the author of this thesis undertook a research project, which incorporates learner corpus data analysis as well as an experimental study. An additional element of the project is a survey, whose aim was to obtain students' feedback on the teaching techniques under examination. The project was divided into several stages and was implemented at Adam Mickiewicz University's Teacher Training College (Kolegium Języków Obcych, henceforth KJO) in Poznań in the years 2009-2011. The first part of this chapter introduces the aims of the study, its subjects, preliminary stages and initial findings; the main body of the chapter presents the quasi-experimental study in detail, followed by the discussion of its outcome.

4.1. Aims

The aims of the research undertaken include employing corpus-based techniques in lessons in advanced English grammar and testing their usefulness and effectiveness in classroom teaching at this level. More precisely, the experimental lessons concentrated on targeting some Polish interference errors which had been observed by the author as persistent in the students' interlanguage, despite her repeated efforts to minimize them. The researcher felt that concordance-based presentation and practice techniques might

be more effective than the more standard techniques used beforehand. All the lessons were part of the third year grammar course at KJO, and the language problems addressed in them had been recognized in the initial stages of the project (error analysis and then pre-test) to be common in the subjects of the study. The techniques employed in the experimental lessons were selected from the variety of corpus-based activities presented and discussed in Chapter Three.

Although the number of language problems needed for the further stages of the study was strictly limited, the items had to be selected from a set of factually occurring, clearly identified problem areas, diagnosed as challenging for students at their advanced level and relevant to their needs. This is why the initial stages required an accurate diagnosis of the interference errors occurring in various types of language use – from strictly controlled situations, with major attention devoted to form (examination essays), to relatively spontaneous and message oriented ones (blog posts). Only items identified in both types of language production were considered for further experimentation procedures.

The process of recognizing these problem areas involves another aspect of corpus-based teaching: the use of learner corpus. Another important aim of the study has been to test its usefulness both as a diagnostic tool and, at a later stage, as a source of teaching materials.

The benefits of the study were intended to be very practical: positive outcomes might encourage teachers to use the new text analysis technologies both for diagnostic purposes and for classroom instruction. Negative outcomes, on the other hand, would indicate that DDL classroom activities should be approached with caution.

4.2. The main hypothesis and research questions

The aims specified above can be now defined for the needs of the research project as a hypothesis and the set of research questions that the study is intended to answer. The main hypothesis posed by the author is that data-driven learning techniques are more effective in eliminating interference errors than the more conventional ELT techniques. This premise is based on the discussion of the learning processes that data-driven learning most likely facilitates, the L1/L2 comparative techniques that DDL can offer, and

the author's own very positive impression of using a corpus for language reference and instruction. Apart from the hypothesis, which is going to be tested in a quantitative study, the study poses three research questions, which can be answered through a less rigid analysis of collected data:

(1) What are common interference errors that advanced Polish learners commit across different levels of formality in English?
(2) Is the effectiveness of corpus-based techniques related to learner-specific features like proficiency, attitude or motivation?
(3) How do students evaluate the usefulness of corpus-based activities?

The first question was approached through two attempts at error analysis – a preliminary one conducted in the traditional 'analog' manner, and the other – corpus-based. Answers to the other two questions were sought by means of a survey conducted among the participants of the experimental part of the project.

4.3. Participants of the study

As stated above, the subjects taking part in the experiment were in their last (third) year of their BA level English studies at Kolegium Języków Obcych in Poznań, which was part of Adam Mickiewicz University and offered studies in English, German and French to young people who plan to become teachers of these languages. This is an important aspect of the study, especially of the survey element, because having received intensive methodology training, the subjects may be assumed to have been more reflective learners – and perhaps more critical – than other advanced EFL students. Generally, they may be said to be fluent speakers of English, some of them actually very proficient. For most of the students, however, the language material of the project lessons was far from obvious, which is demonstrated by the results of the study.

This researcher was not involved in defining the groups' setup, which is why there was no prior knowledge considering their composition, level of English or any other factors which could have affected the results of the experiment. Since the assignment of participants to the experimental and control groups was not totally random, the study must be classified as quasi-experimental rather than purely experimental.

For reasons presented in detail below, there were two experimental attempts carried out for this project, and so two sets of students were involved, one in the academic year 2009/2010 and the other, for the final version, a year later.

4.3.1. Academic year 2009/2010

At the beginning of their studies (in year I) the students were divided into three groups on the basis of the English accent they wanted to speak (two British English groups – randomly formed – and one American English group). In the third year of their studies, though, when this study was carried out, the groups were rearranged into two for some classes, and that was the case for the grammar course as well. They will be referred to as group A and B. The actual grouping was affected by many various factors, such as the schedule for other classes students were supposed to attend or their original group membership (accent). There was not one key according to which individual students were assigned to the groups.

Another characteristic of the groups was a fluctuation in their size. In the winter semester there were 14 students in Group A and 15 students in group B. In the summer semester, however, a large group of students returned from their Erasmus scholarships. Most of them joined group B, which had 21 members as a result. Group A grew only slightly, and consisted of 16 students. What complicated the situation even more was the fact that the project was extended over time (six lessons over two semesters), so inevitably attendance varied, with some students absent from some lessons. As a result, the number of participants of the project had to be reduced to those who attended most of the lessons – an exception had to be made for single absences in the sequence of the six classes, or there would be hardly any subjects available. In order to minimize the effects of the absences, the teacher supplied every student who missed a class with teaching materials and relevant instructions. Thus the number of students involved in the project from beginning to end was very small: in the smaller group (A) it was 11 students. Consequently, to make comparisons between the groups possible, the other one was also reduced to that number. For the sake of balance the same number of students were selected from among those students from group B who attended the classes in a similar pattern (4 students with a full attendance record and the others with single

absences allowed). The choices were also motivated by another factor, expounded below.

The grouping of subjects for the study was challenged by differences in their level of proficiency in English. All measurements applied to the two groups indicated differences of too large proportions to be ignored in the research. Following procedures recommended by Dörnyei (2007), the results of a few weaker students were deleted from the group A data and the strongest students' results were excluded from the group B data. Details of this manipulation are presented in section 4.9. , p. 242. As it is going to be evidenced, the differences between the groups were much deeper than their proficiency level, and the manipulation described above failed to eradicate this disparity.

4.3.2. Academic year 2010/2011

In the academic year 2010/2011, the students were divided into two research groups as well. One combined 22 students from two separate groups (starting with 11 students in each), while the other consisted of 19 students (total - 41). As can be expected in studies performed in educational institutions, a few cases of student attrition occurred, with the result that at the end of the project the total of the students was 39, with three students who abandoned the course and one who returned from an Erasmus visit for the summer semester. The sizes were relatively equal, then, and, as it later transpired, the level of proficiency did not differ significantly in the two groups. It was not possible to establish this beforehand: many of the students had no Practical English examination results available because they had spent the previous semester on their Erasmus visits abroad, and they took their English exams at their host universities. In the grammar course, however, the students took several grammar tests whose combined averages differed insignificantly between the two groups: 67.88% and 66.35%. The difference was verified in an independent samples *t*-test assuming equal variance (verified with an F-test), with the *p*-value of 0.64, which failed to reject the null hypothesis of there being no significant difference between the two groups' average test results. The students' attitudes towards the corpus-based techniques also seemed more equally distributed than the year before, each group holding both enthusiasts and opponents of these. Survey results discussed near the end of this chapter will cast more light on this issue.

The participants of the study were informed at the beginning of the course that part of it would be devoted to eliminating their common Polish interference errors, and they accepted this idea thoroughly. They were also told that a research project would be conducted during the course, although little detail was revealed as to the nature and subject matter of the study in case that affected their performance in the subsequent tests.

4.4. Research plan and procedures

The project was initiated in 2009, although the student-generated language material had been collected since as early as 2006. The study had several stages of development:

(1) collecting and analyzing interference errors in examination essays;
(2) building a corpus out of students' blog entries in English;
(3) searching the corpus to see which of the errors encountered in the exam papers would reappear in the less controlled and more informal context of blog writing;
(4) the experiment – the first stage (2009/2010)
    a. choosing items for the experiment – aspects of English identified as interference error-prone in both formal and informal styles;
    b. teaching the error-prone language problems in experimental lessons with the use of corpus-based materials and, in the control group, by employing other forms of language instruction;
    c. administering a post-test to compare the effectiveness of the two procedures;
    d. conducting and analyzing an anonymous survey with the view of receiving feedback from students on corpus-based activities;
(5) the experiment – the second stage (2010/2011)
    a. choosing items for the experiment – aspects of English identified as interference error-prone in both formal and informal styles;
    b. designing and applying a pre-test for the items selected;
    c. teaching the error-prone language problems in experimental lessons with the use of corpus-based materials and, in the control group, by employing other forms of language instruction;

d. administering a post-test to compare the effectiveness of the two procedures;
e. administering a delayed test to observe the duration of the effect established in the post-test;

(6) comparing the pre-test/post-test gains of the experimental and control groups for particular language items;
(7) conducting a non-anonymous survey with the view of receiving feedback from students on the corpus-based activities which they had done in class;
(8) analyzing survey results and trying to establish possible correlations with test statistics.

4.5. Research instruments

The experimental part of the study involved a battery of instruments, carefully selected to appropriately address the questions posed at the beginning of the project. The starting point of the process was error analysis – first manual, for which a database was built in *Microsoft Access 2007*, and then corpus based, utilizing the *TextSTAT* text analysis software developed by Hünig (2008). These steps were taken to select target items for further testing. The next stage was the pre-test, a set of Polish sentences containing several instances of target items, to be translated into English. Students' translations of the test items were assessed, and the reliability of the test was then verified by means of the Spearman-Brown split-half estimate (Bachman 1990, Bachman 2004). The aim of the pre-test/post-test design was to measure the progress that students made in their use of the items addressed in the lessons. The statistical instrument that had been planned for the purpose was an independent samples t-test, performed on gain values for individual students. One of the key requirements of any t-test, however, is the normal distribution of the results, which was measured for test results on each of the six language points separately by means of the Shapiro-Wilk test of normality. For most lessons this condition had not been met, which is why a non-parametric equivalent of the independent samples t-test, the Mann-Whitney U test had to be applied (Brown 1988). The term 'gain' refers to the difference between the number of errors a student committed before and after the treatment. This is the dependent variable in the experiment. The bigger the

gain was, the more successful the lesson was considered to be. The statistical test was to verify whether the experimental, corpus-based treatment was indeed more effective in this respect than other, more conventional classroom activities. A paired-samples test, often used in effectiveness studies, would not be sufficient if any comparisons were to be made, but would only confirm or disconfirm the effectiveness of the experimental treatment. Such a test was, however, applied to analyze the results of the post-test and the delayed test in order to examine the effect of the time factor.

The majority of statistical calculations for the project were performed with the use of *IBM SPSS Statistics*, though some preliminary analysis was carried out with the *Excel* data analysis add-in. Most results obtained in the experimental part of the study are presented in box-and-whiskers plots, or boxplots, which illustrate some of the important facts revealed in the statistical analysis of the test results. The upper and lower bounds of the boxes represent Tukey's hinges, i.e. the range of 25 percent of results (a quartile) upwards and downwards from the median, which is itself marked with the thick line across the box. The length of the box represents the interquartile range (IQR) – the distance between the hinges, which covers the central 50% of cases. Values between 1.5 IQRs and 3 IQRs from the edge of the box are labeled as outliers (○). Values of more than three IQRs from the edge are labeled as extreme, and denoted with an asterisk (✳). The "whiskers" indicate the highest and the lowest observed results that are not outliers, i.e. are within 1.5 IQRs of the hinge. The more symmetrical a boxplot is, the more normal is the distribution of data that it illustrates. Other graphs placed in the thesis include frequency bar charts, radar graphs, histograms and pie-charts. The last two were all generated with *SPSS* software, as were all the boxplots.

Results for each set of three lessons of the same type (corpus vs. non-corpus) in the same group of students are analyzed by means of the independent samples Kruskal-Wallis H test, which is a non-parametric equivalent of the one-way ANOVA. This was done in order to ascertain whether there were any significant differences in the effectiveness of different items within a given set. See section 4.10.4.2. for more information. Analysis of the data obtained from the experiment ends with a broad comparison of pooled results for the corpus-based lessons with those for the conventional lessons in each group.

In order to triangulate the findings of the experiment, the author introduced another instrument into the project: a survey. Students were asked to express their opinion

on the experience of DDL lessons, and specifically on the techniques applied in them as compared with those they were more familiar with. Participants were also required to provide some information on their background and plans for the future. The data gathered in the survey were used in the correlational analysis, which is the last component of the study. Wherever distribution of data allowed (i.e. did not depart from normal), Pearson's correlation coefficient (*r*) was calculated. For variables whose values did not prove parametric, Spearman's rank correlation coefficient (ρ) was used.

The rest of this chapter presents each of the stages, with analyses of their results, followed by a summary of the findings of the study and conclusions.

4.6. Analysis of interference errors in examination essays

As stated earlier, one of the key aims of the study was to identify interference problems common in Polish advanced learners' English. Then several such errors were selected to constitute the subject matter of a sequence of special remedial sessions in the third year grammar course at KJO. The subjects' explicit knowledge of English ("conscious awareness of minor and major schemas" – Ellis and Barkhuizen 2005: 5) can be accessed through analyzing forms which do not conform to the rules of English grammar. For the purpose of such an analysis a database of interference errors was built on the basis of the end-of-year examination essays of the students involved in the experiment. The assumption here was that these data represent students' conscious knowledge of the rules of English grammar and lexis, as the examination was carried out in highly controlled conditions, and with very high stakes involved (the examination must be passed for a student to be admitted to the final year of his/her studies). Therefore the students are believed to have applied the highest level of monitoring of their output, though restricted by the time limit assigned for the task (400-450 words in 120 minutes). It was important that the authorship of the essays be unquestionable and the writers depend solely on their command of English rather than reference materials or other people's support. Such conditions were, indeed, met in the practical English examination essays.

The exam for the purposes of the first analysis took place in July 2009, when the students were just finishing their second year at the college. These students, as well as their colleagues the following year, were later to participate in experimental lessons, so

it was important that their particular problems be identified and addressed. There were 33 students taking the exam, and they had three topics to choose from: one concerned the advantages and drawbacks of lowering the school enrolment age in Polish schools, another was an article on the moral and professional standards of today's media, and the third was a letter to the Olympic Committee in which the student was to express his/her opinion on the role of high profile athletes in society. The most popular topic (19 essays) was the first one. This is quite understandable, because KJO students have (naturally) both an interest in and quite good knowledge of issues connected with education. The second was chosen by six students, and the third by eight.

The analysis of these papers consisted in identifying errors which *may* have arisen due to the influence of Polish, the subjects' L1, and then in classifying them according to the criteria discussed in Chapter One. The errors were collected into a database, which made it easier to find particular types of errors at later stages of the analysis. Each error was labeled with a student code, its individual number and the categories it had been assigned to. Also, each entry included a "plausible interpretation" (Corder [1973] 1981), in which the researcher attempted to reconstruct the intended meaning and appropriate wording of the erroneous fragment. The focus being on transfer errors here, every error was accompanied by the assumed L1 sources of the error. Although it was impossible to state with full confidence how a learner had arrived at a particular interlanguage form, the errors listed do demonstrate features of the student's mother tongue, which may indicate at least a partial role of L1 in causing the deviation from the L2 norm.

In agreement with recommendations from Ellis and Barkhuizen (2005), the labeling system was devised for the language material obtained in this study, so as to accommodate the collected data most adequately; in other words, the system was not intended to constitute a universal classification of interlingual errors. On the other hand, the categories applied are relatively broad, so they are flexible enough to embrace a rich variety of language data. Devising a classification scheme with a very large number of levels, categories and sub-categories may lead to an excessively complicated system, which causes confusion instead of facilitating a comprehensive and clear analysis. What has been achieved here is a relatively straightforward organizing method, sufficient for the purpose of this study.

In consideration of what is emphasized in Dulay et al. (1982) as a potential problem in error analysis, namely, the lack of distinction between analyzing *product* (the description of an error) and *process* (the explanation of an error), care was taken to devise a system that would keep these two aspects of analysis separate. For this reason there are two sets of categories: descriptive (primary) and explanatory (secondary).

4.6.1. Descriptive taxonomy

Around 300 errors which had been collected (see Appendix A) were divided into three broad descriptive categories identified by the aspect of the language system affected: errors of form, lexico-grammatical errors and syntactic errors. This set of classes with their sub-categories seemed most pedagogically relevant to the aims and context of this study. The experimental lessons were to be part of a grammar course, so the items selected for them needed to be merged into the grammar syllabus. Such classification of the data made this possible. Some overlap was observed between this broad classification and Lennon's (1990) idea of *domain*, the linguistic unit required to identify the error: for errors of form, the domain is most often *word* or *phrase*, lexico-grammatical errors are most often analyzed at *phrase* or *clause* level, and the domain for syntactic errors is mostly *sentence* or *clause*. The domain of *text* does not seem to have any regular pattern and appears wherever the identification of an error is linked with the style or the general message of a particular text. The basic set of categories and sub-categories is as follows:

(1) Errors of form:
    a. grapho-phonemic
    b. inflectional
    c. word-formation
    d. categorial
    e. formal
(2) Lexico-grammatical errors
    a. agreement
    b. article omission

c. colligation
d. collocation
e. stylistic errors
f. tense/aspect
g. word choice

(3) Syntactic errors

a. clause combining
b. clause combining/punctuation
c. clause structure
d. parallel structures
e. voice
f. word order

The following section defines each of the categories and discussess their characteristics.

4.6.1.1. Errors of form

Errors of form are those errors in which learners use deviant, inappropriate or non-existent forms of English words, or violate conventions of written English:

- grapho-phonemic errors – most of these are wrong spelling resulting from different relations between written and spoken forms in L1 (e.g. *generaly, *intelectual, *clons). This group also incudes problems with other conventions of written English, especially rules for writing numbers and numerals (e.g. single-digit numbers written out as words rather than given in figures);
- inflectional – despite the very limited inflection system in English, Polish learners seem to struggle with constructing and using one particular form: the Saxon genitive. Although the problem may not be unique to Polish learners and is a cause for concern to English teachers worldwide, the fact is that inflection involving punctuation marks is a rare feature among languages and that is why it causes frequent problems to learners of English as a foreign language of various nationalities;
- word-formation – non-existent forms resulting from compensatory strategies, when students resort to derivation on the basis of analyzing L1 words into morphemes and

translating these into English (*biligual transfer* in Lewandowska-Tomaszczyk 2001: 227);

- categorial – the use of language items of similar meaning but from a wrong grammatical category (here: determiner vs. pronoun).

4.6.1.2. Lexico-grammatical errors

Lexico-grammatical errors constitute the biggest group, encompassing lexical and grammatical problems, mostly at the level of the phrase. It became evident when the error analysis was being performed that it is very artificial, if not virtually impossible, to keep these two major aspects of language separate. Lexical choices are limited by context, by the internal logical structure of the text, the clauses or the phrases in which words appear, and these need to be considered in attempts to reconstruct the student's intended utterances. Conversely, words have their local grammar, i.e. they need to be surrounded by particular forms in strictly defined structures, and when the learner fails to provide them, it would be very artificial to say it is a purely grammatical fault. The rules of grammar are in most cases inseparably linked to the lexical choices language users make. This is in accordance with Sinclair's (1991) concept of *lexicogrammar* and with Lewis's (2000) recommendation to use it as one broad field of language analysis, rather than lexis and grammar separately. The concept has been broadly discussed in Chapter Two, section 2.2. The following subcategories have been identified for the lexicogrammatical errors:

- agreement – errors resulting from the lack of concord between subject and predicate or pronoun and antecedent[25] (e.g. *real money that ENTER national budget)
- article omission – a separate category in this study because of the high frequency of articles in English on the one hand, and high frequency of article omission errors in

[25] There have been some doubts whether this group should belong with lexico-grammatical errors or with the previous group, as agreement is a problem of form. However, since this is a context-dependent error, determined by characteristics of other words occurring in the phrase (e.g. *the majority of their memories IS...), it was classified as a lexico-grammatical problem.

the data on the other; Polish learners' tendency to commit such errors was a subject of discussions on transfer among language acquisition specialists:

> Arabski (1979) made the somewhat surprising assertion that the 974 article errors in his Polish-English corpus were not transfer errors on the grounds that, because Polish does not have articles, there is nothing to transfer. Clearly, though, the absence of a structural feature in the L1 may have as much impact on the L2 as the presence of a different feature (Ellis 1994: 311-312).

It must be said that Arabski later changed his position on the issue in Arabski (1990), where he says that the "difficulty [of articles] is caused by the structural difference between Polish and English and it is the result of transfer". More studies of learner performance seem to confirm that speakers of languages which do have articles use them more accurately in English (e.g. Oller and Redding 1971). Dušková (1969) recognized frequent article omission in her Czech students' English as transfer errors, on the grounds that their L1 had no clear equivalent. Other article errors (article misuse or overuse) were recognized to be intralingual errors, resulting from the complexities of the reference system in English. This is the stance assumed in this study as well.

- colligation – originally defined by Firth (1957) as a purely grammatical concept (the co-occurrence of grammatical categories) but later reformulated by Sinclair (1998) and Hoey (2005) to mean a relation between a node word and grammatical categories which frequently co-occur with it. This is how the concept is now used in corpus-driven language studies. Lewis gives a similar definition (2000: 137): "[C]olligation is the way one word regularly co-occurs with a particular (grammar) pattern so, for example some verbs typically occur with a particular tense, or a noun might typically appear preceded by a personal pronoun, rather than an article". Such tendencies concerning surrounding language are particularly frequent in English nouns and verbs: they allow a restricted set of forms of possible complementation (noun / infinitive / gerund / clause / prepositional phrase), or tend to occur in some forms but not in others (e.g. *have* in its possessive meaning is not normally used in the progressive aspect). This is a frequent source of learners' problems;
- collocation – Lewis (2000: 132) defines collocation as "the way in which words co-occur in natural text in statistically significant ways". Therefore, collocational errors involve use of words which do not co-occur, giving the reader/hearer the impression

of awkwardness and unnaturalness (e.g. **throw arguments*, **stressful factor*, **to stand the law*).

- stylistic errors – a rather small group in the sample, but important to distinguish: stylistic errors involve the use of words which clash with the style (or register) of the discourse they appear in. Their L1 source is usually the absence of stylistic variants among L1 equivalents of a particular lexical item (underdifferentiation). Example: *interaction with other HUMANS* (scientific register);
- tense/aspect – tense and aspect systems are very different in English and Polish, emphasizing different aspects of meanings connected with time. "In Polish the aspectual distinctions are realized on the lexical level by lexical means such as prefixation, suffixation and suppletion. (...) [T]here is only one type of aspectual opposition: perfective versus imperfective. With a few exceptions each Polish verb is either perfective or imperfective" (Fisiak et al. 1978: 107). English, on the other hand, has two aspectual dimensions – perfective and progressive, and there is the third, unmarked option, absent from the Polish system. These differences between the tense systems are at the root of some tense and aspect errors in Polish learners' English.
- word choice errors – the difference between this category and the one above is that the problem lies in the choice of a particular lexical item to express a given meaning, rather than making a word fit smoothly other items surrounding it. The context is still important in determining the meaning of the utterance, but there does not need to be a particularly strong collocation involved.

4.6.1.3. Syntactic errors

The category of syntactic errors comprises errors of structure which are not connected with lexical choices, but are related to clause or sentence structure. These include:

- clause combining – a big group of errors involving problems with linking clauses into appropriately structured sentences (e.g. *LIKE Phelps did*, use of *however* instead of *although*);

- clause combining/punctuation – errors of punctuation related to combining clauses and resulting from different rules in equivalent types of clauses in Polish (mainly use of a comma with noun clauses and relative clauses);
- clause structure – clause elements erroneously put together or simply missing; most errors in this group are subordinate subject omissions and problems with negation;
- parallel structures – errors involving the faulty construction of parallel phrases (e.g. **separating what is true from FALSE*);
- voice – problems connected with an inappropriate choice of voice (passive or active) or faulty structure thereof;
- word order – problems with ordering elements in the sentence (e.g. emphatic inversion, participial adjectives, adverbial placement, viewpoint adverbs).

4.6.2. Explanatory taxonomy

The last column in the database offers a plausible explanation of each error and its likely L1 sources, which may be wholly or only partially responsible for the occurrence of the error. The observed parallels between the learner-produced form and his/her L1 may indicate that there is a causal relationship involved. Those relationships are indicated by labels which are placed at the beginning of each entry:

- underdifferentiation – occurs when a particular L1 form has two different counterparts in the TL. This concept was initially introduced by Weinreich ([1953] 1968) with reference to phonological features in bilinguals, and later adopted by other SLA researchers. For example, Arabski (1979) discusses the phenomenon at length in reference to lexical errors. In this study the category is not limited to lexis. *Underdifferentiation* errors are observed across the board, among formal, lexico-grammatical and syntactic errors.
- calque – errors which are word by word translations of L1 phrases into L2;
- coinage – a parallel category to that of calque, but operating at the morphological level: errors resulting from a morpheme by morpheme translation of L1 words into L2;
- feature absent from L1 – errors resulting from the fact that a particular target language feature (lexical or morphological element, semantic distinction, grammatical

or syntactic rule, category or distinction) does not exist in L1, causing difficulty in using such features appropriately. In this data collection the most common errors resulting from such situations are article omissions and omissions in general, but other surface strategies do appear as well;

- L1 feature absent from L2, which results in application of rules not actually operative in L2: the material evidenced numerous orthographic errors which may be connected with the fact that doubled consonant spelling in Polish is associated with gemination in speech, which is not the case in English. Hence, problems with double-letter spellings in words like *inappropriate*, *partially* or *intellectual* could have been aggravated by L1 features;
- "false friends" or deceptive cognates – errors resulting from formal similarity between L1 and L2 forms, whose distributions and/or meanings are different;
- transfer of training (Selinker 1972) – errors which originate from the way the learner was taught. This normally refers to intralingual errors, but may sometimes include L1 formal instruction affecting L2 production, e.g. punctuation rules in Polish being transferred into English because of their prominence in the Polish language education;
- foreignizing – a communicative strategy in which the learner adapts L1 linguistic items to L2 grapho-phonemic system in the hope that such a word exists and/or will facilitate successful communication (Faerch and Kasper 1983).

4.6.3. Comments on essay errors

The data set collected for the study showed a variety of issues with which the learners struggle. The biggest number of interlingual errors observed for one learner was 29 (code 2005) and the smallest was 5 (code 2020). On the whole, the analysis of the essay errors was not quantitative, but aimed at identifying the types and the level of problems students had at that stage of their L2 development. It was also important to find out in what ways L1 affected the learners' interlanguage. Groups of errors strongly represented in the data are discussed below.

4.6.3.1. Errors of form – 49 items

As already stated, errors of form include items which do not exist in English, violating the conventions of written English. The group can be divided into the following sub-categories:

- grapho-phonemic errors;
- inflectional errors;
- word-formation errors;
- categorial errors.

Grapho-phonemic errors: In the essays, the most common spelling errors involved learners' failure to double letters where necessary (**aford*, **intelectual*, **willingnes*), especially with the adverb-forming suffix {-ly} (**totaly*, **publicaly*, **partialy*). It may be connected with the fact that in standard Polish, where consonant letters are doubled, gemination occurs, i.e. the consonant is lengthened (*lekki* vs. *leki*). Hence frequent spelling errors in words with doubled consonant letters, because the L1 phonetic feature associated with this spelling (gemination) does not occur in English. The opposite may happen: learners may remember that some letters are doubled in a given word, but have no phonetic clues which ones they were, and this is the source of spellings like **papparrazi*. Another source of errors is reduced or silent vowel letters (**existance*, **immensly*, **clons*, **easly* or **modal* –for *model*). There were also spelling problems with letters representing vowel distinctions and vowel sounds that are particularly distant from the Polish phonological system: **stuff* (for *staff*) or **vertues* (for *virtues*).

One error seems especially interesting from the IL point of view: the form *price* instead of *prize*, which must result from final devoicing, a very prominent feature of the Polish accent in English: "Word-final obstruent devoicing is probably the most notorious characteristic of Polglish, and predictably so, since this is a universal phonological process reinforced in Polish speakers by the system-adequacy" (Dziubalska-Kołaczyk et al. 2006: 2).

As many as 13 out of 33 students in the group had L1-based problems with orthography, which may be surprising, considering that in the written exam situation the monitor for this aspect of language is usually quite high.

Inflectional errors: As far as inflection is concerned, there is one form that causes Polish learners some problems: the Saxon genitive. The source of the problem may be underdifferentiation, as English has two forms equivalent to the Polish genitive: the Saxon genitive and the periphrastic genitive (X of Y). These forms are not in free variation, however, the former being preferred with human and animate nouns, the latter with inanimate nouns as 'possessors'. This causes confusion and may be a source of misselection errors (choosing one form instead of the other). What makes matters worse is the fact that Polish does not use punctuation for marking inflection, which may add to the difficulty learners encounter with using the Saxon genitive. Other errors observed are less grave, and include omission of the Saxon genitive apostrophe in plural NPs, and finally one case of the Saxon genitive being used with a generic adjective (*the mature's), which English grammar does not allow.

Also among the grapho-phonemic errors are problems connected with using numbers in formal English texts. Polish learners find it hard to remember (or do not realize at all) that English formal style requires single digit numbers to be written out as words, and this causes frequent stylistic problems for Polish users of written English. Equivalent rules for Polish are less definite and are rather considered part of specialist knowledge (cf. Wolański 2008).

Word-formation: learners sometimes resort to coinage, a common communication strategy, trusting that their knowledge of L2 derivational morphemes will allow them to produce a valid and understandable L2 lexical unit. This certainly is not always the case; hence errors like **cruelity*, **unharmful* or **publishment*. These are identified as errors of form rather than lexico-grammatical, as they are non-existent forms in the English language. They may still be communicatively successful, but are definitely recognized as foreign to the English language system. Probably a more hazardous strategy is *foreignizing*, where a learner applies some L2 features to an L1 lexical element. The few instances of this strategy in these data are **horizont*, **psychologic* and **dotation*.

4.6.3.2. Lexico-grammatical errors – 191 items

As pointed out above, the lexico-grammatical group of errors is the biggest in this collection, which is not surprising considering the breadth of the category itself. It is main-

ly colligation errors that were behind choosing the category of *lexico-grammatical errors* for this study, instead of keeping grammar errors and lexical errors separate. Such was the assumption before the analysis was undertaken, and the process itself confirmed that keeping grammar and lexis separate in error analysis, especially at the advanced level, would have been impossible. Each of the sub-categories is discussed separately below. These include:

- agreement
- article omission
- colligation
- collocation
- stylistic errors
- tense/aspect
- word choice

Agreement errors: Significant differences between the ways in which Polish and English behave concerning concord cause quite a large number of errors even in advanced learners' English. A typical error of agreement involves a subject noun which is uncountable in English and plural in Polish, which results in a choice of the English verb form that agrees with the Polish plural concept in the subject of the sentence (e.g. **money RULE the world*, **information that sometimes ARE not worth seeing*). What is more, number, a grammatical feature generally familiar to speakers of Polish, may be confusing in some English words whose form is identical with the regular plural (suffix {-s}), but which are in fact uncountable (e.g. *news*, *measles*, *mathematics* – the last one either singular or plural, depending on meaning).

An interesting case of agreement error is the following utterance: **the number of births DECREASE every year*. The error is possibly induced by classroom instruction connected with the use of the word *number* in English, quite different in terms of agreement from its Polish counterpart *ilość*, which is always used with singular verb forms. Polish students find it difficult to distinguish between the following two uses of *number of*:

The number of births DECREASES every year.
singular verb form – concord with *number*

‘Ilość urodzeń MALEJE co roku.’

singular verb form – concord with *ilość* (here: *number*)

A number of births WERE unattended by a doctor.

plural verb form – concord with *births*

‘Pewna ilość urodzeń ODBYŁA się bez opieki lekarza.’

singular verb form – concord with *ilość* (here: *number*)

An attempt to avoid an error of this type is a very likely source of the observed error. One may guess the student who made the error was aware of some difficulty the word poses in terms of agreement, but failed to remember (or apply) the rule correctly. The same problem appears with the use of the word *majority* and its equivalent *większość*.

Some problems with pronoun-antecedent agreement were observed as well, especially where reflexive forms were involved. This may be attributed to the fact that Polish has only one lexical item for the emphatic reflexive (*sam*) and for the possessive reflexive (*swój*). These are inflected as all nominal forms in Polish are, but they do not vary lexically with the grammatical person and number as they do in English. This may be the source of special difficulties students encounter with reflexive forms in English.

Article omission errors: As said above, these errors are very common for Polish learners of English as well as for learners whose L1 is Czech, which, like Polish, does not have any systemic equivalent to articles. The difficulty lies in the multiple grammatical distinctions which need to be made for the article to be used, and used correctly. First of all, the learner needs to *remember* that the decision concerning article choice must be made at all. Second, s/he needs to decide whether the reference is specific or non-specific – admittedly a very abstract distinction. If reference is non-specific, other factors come into play: countability and number. All this needs to be done without any support from L1, where none of these distinctions are marked. It is not surprising, therefore, that article omission features so prominently in the data set: 45 out of all the 298 errors were omissions of either the definite or indefinite article, i.e. 15%. It is even more impressive if the number of students who had such problems is considered: 18 learners out of 33, which makes 54.6% of the group.

Colligation errors: Most of the relatively few (14) errors in this group have been attributed to calque – the literal translation of L1 phrases into L2. This strategy is what

learners resort to if they do not know colligation restrictions for particular lexical items: they seem to follow the patterns of their L1. This is the most probable source of errors of choice between the infinitive and gerund, and of the misselection or omission of prepositions.

Collocation errors: This is another big group observed in the data – 54 items. The most common factor involved in generating the erroneous forms seems to be calque. A variety of word class combinations are involved, the most common being dependent prepositions (**enroll a child TO school*), VERB + noun (**STAND the law*, **ACQUIRE good results*), ADJECTIVE + noun (**a PEJORATIVE feature*), adjective + NOUN (**moral DISASTER*) and prepositional phrase (**IN the age of six*).

Word choice errors: There are 50 word choice errors in the data – most often single misapplied words (e.g. **ECONOMICAL crisis*, **sweet CARELESSNESS* or **JOIN classes with play*), most of which seem to have been chosen as a result of underdifferentiation. In these cases there is usually one Polish word which may have several equivalents in English, but the one selected by the student has a different meaning (or function) in English than s/he is assumed to intend. Hence errors involving the choice between *possibility* and *opportunity*, *alter* vs. *change*, *learn* vs. *study*, or – more grammatically oriented – *some* vs. *a* in front of singular countable nouns. The last example has a semantic component to it too, as *some* used in front of singular countable nouns is marked (strong form) and carries special, usually negative meaning (Swan 1995). Learners have little awareness of this and use the form instead of a marker of indefiniteness, i.e. an indefinite article.

Apart from lexical errors involving single words, the word choice error category includes wrongly built multi-word phrases as well, because their intended meaning may be indecipherable to a native speaker of English (e.g. *experimental rabbit* or *hurry with the material*). These most often result from loan translation, i.e. calque.

An interesting phenomenon features in the data in connection with the word *careless*, which in English is an antonym of *careful* and has a clearly negative prosody. From the context of the essays about childhood and the reasons why children should or should not be sent to school earlier, the reader familiar with the students' L1 can deduce that the intended word was *carefree*. As many as five participants of the study seem to have coined the former (*care+less*) with the intention of a clearly positive semantic prosody by analogy from Polish ('bez+troski'), with the same derivational pattern as the

one involved in such words as *thoughtless*, *painless* or *hopeless* (Polish: 'bez+myślny', 'bez+bolesny', 'bez+nadziejny'). An intralingual error may be involved here as well, as an overgeneralization of the adjectival suffix {-less}. What is the most likely, however, is that the error is a result of both interlingual and intralingual processes. Some of the five learners went even further than that and nominalized the thus coined *careless* by adding a nominal suffix {–ness}, producing *carelessness*, where the intention must have been *lightheartedness*, though admittedly a much more difficult word to remember and/or construct.

Stylistic errors: Not many stylistic errors can be associated with interlingual influence. In the data collected, only six such errors were recorded, most of them resulting from underdifferentiation. Students do not seem to be aware of the stylistic clashes resulting from word combinations like **stimulate mourning* (instead of *cause resentment*) or *interactions with other humans* (possible in anthropological or philosophical texts, but rather awkward in a general academic essay). The latter is perhaps less drastic but does show lack of stylistic sensitivity: *interactions with other people* would be much more appropriate here.

Tense/aspect errors: There are not many tense and aspect errors here, probably because of the type of discourse students produced. Most texts are expository or argumentative, and these modes of writing do not usually have complex temporal references. Another explanation could be that with the tense and aspect systems of Polish and English being very different (Fisiak et al. 1978: 96ff.), a lot of attention is paid to the problem in English language classes in Poland. The low number of such errors in this data set may partly be a result of successful teaching.

The errors that do occur are observed in more complex verb phrases, often involving infinitives or the hypothetical past, which in Polish is expressed in very different ways. There is no grammatical distinction between present and past hypothetical meanings and the verb has a unique inflectional suffix {–by} for expressing them, rather than a series of auxiliaries preceding the lexical verb:

He would not have made a career. – 'Nie zrobiłby kariery.'
He would not make a career. – 'Nie zrobiłby kariery.'

4.6.3.3. Syntactic errors – 51 items

Polish and English are very different language systems, and those differences are probably most evident in syntax. In Sapir's terms (Sapir 1921: 74) the former is predominantly synthetic (a lot of grammatical information is accumulated within one word and its selected form) while the latter – analytic (grammatical information is mostly provided by free morphemes). Generally, these many discrepancies between the two syntactic systems cause Polish learners of English numerous problems, which can be classified as follows:

- clause combining;
- clause combining/punctuation;
- clause structure;
- parallel structures;
- voice;
- word order.

Clause combining: These errors are evidence of some students' struggles to build grammatical compound, complex or compound-complex sentences. Frequent problems involve choosing an appropriate subordinating conjunction (e.g. the preposition *like* often used instead of *as* in comparative clauses, or **the way how* instead of *the way that* in relative clauses). For the needs of this study some errors of clause combining are given an additional label (*punctuation*) because some punctuation errors are strongly connected with problems at the syntactic level. One such error is a comma splice, where two independent clauses are held in one sentence without a conjunction that would link them, with a comma instead. Often there is an attempt to link those independent clauses, but the connector is wrongly selected: instead of a conjunction (e.g. *although* or *but*), learners choose linking adverbs (e.g. *however*), whose grammatical function is not to link clauses structurally but to show logical relations between separate sentences or clauses linked by other grammatical devices (e.g. *thus* used with a participle clause). That such distinctions are not very clearly defined in Polish can be seen in examples from the National Corpus of Polish (Pęzik 2012):

> *Horoskopy to fajna sprawa, <u>w końcu</u> każdy w coś wierzy i czytając ma nadzieję, że to się spełni.*

‘Horoscopes are great; after all, everybody believes in something and, when reading them, they hope that the horoscopes will come true.’

Wypler wynalazł własną metodę nauki pisma chińskiego, niestety, zaledwie okruchy jego dorobku zostały opublikowane.
‘Wypler invented his own way of learning how to write Chinese; unfortunately, barely scraps of his work were published.’

Petrarka zaczął pisać Secretum *w porze, która ciągle jest przedmiotem sporów, wiadomo też, że wracał do tekstu i przerabiał go w ciągu kilku lat.*
‘Petrarch began writing *Secretum* at the time which is still a bone of contention; he is also known to have returned to the text and rewritten it over a few years.’

When translated into English, these sentences need a different structure and punctuation: the linking or viewpoint adverbs showing the logical connection between the two clauses need to be preceded by a semicolon, or can even begin a new sentence. If the Polish punctuation was preserved, the resulting translation would be a typical comma splice – a sentence which consists of two independent clause structures which are not connected by any subordinating or coordinating device. Classifying these errors as merely problems of form (where punctuation normally belongs) would not account for students’ problems with syntax demonstrated in sentences of this kind: difficulty in recognizing what does and does not constitute a dependent and independent clause and in combining clauses into larger structures. Hence the dual labeling for such errors (clause combining/punctuation).

Another typically Polish error of clause combining is the use of a comma in front of the conjunction *that*, which is not normally done in English unless there is an embedded phrase in front. Although there is no evidence of surface *syntactic* fault here, these errors are placed in the syntactic group because they are strongly connected with using particular types of clauses (noun clauses and relative clauses). The error is very likely to be connected with the rules in Polish which require a comma to be used both in front of *że*, the Polish equivalent of *that* in noun clauses, and in front of *który*/*która*/*które* – equivalent to *that* in relative clauses. In this study the error is accounted for as a transfer of training, because this punctuation rule is very explicitly and

persistently taught by teachers of Polish in primary and secondary schools, which may make it particularly difficult for learners to accept (and remember) that English rules are exactly opposite in this case. As the corpus data analysis below will demonstrate, the problem is even more complex: there are other L1 sources of *comma + that* errors in Polish learners' English.

It is worth pointing out that the comma causes problems with the comprehension of such sentences, which in turn makes students' written communication less successful. This, as well as the relatively high frequency of the problem in the data, is an indication that the use (and non-use) of commas with *that* should be dealt with in class more effectively, and – ideally – at earlier stages of English instruction. Further stages of this study revealed other facts about this particular error too (see section 4.10.1.).

Clause structure errors: Most faults in clause structure identified in the material involve the omission of elements required by a structure the student attempted to build. Most often these are missing subjects in subordinate clauses or the omission of an object. The first seems to be deeply grounded in Polish syntax, where subjects are very often omitted and have no lexical presence in the sentence. Hence sentences like this are typical cases of calque from Polish: **The child is better prepared … because Ø has started earlier.* Other errors of this type are also word-by-word translations of sentences in Polish, which, because of the rich inflection system reflecting the semantic roles of nominal elements in a clause, allows such omissions.

Word order errors: Word order rules are much more flexible in Polish, although English word order is not as fixed as it is usually believed to be (Fisiak et al. 1978). Still, there are many limitations on how information should be organized in the English sentence which are absent from Polish. These discrepancies often give rise to word order errors. Thus, the most common word order problems in the data set occur in structures like negative adverbial fronting inversion (e. **Not only they receive...*), adverbial placement (e.g. **keep all the time up to date*) and noun pre- and post-modification (e.g. **mentioned above aspects*).

The material discussed above was the starting point for establishing the real needs of KJO students in terms of remedial teaching. Several areas of difficulty emerged from this analysis: doubled-letter spelling, clause combining and punctuation, agreement, dealing with uncountable nouns and a variety of lexical choice errors. These were investigated in further stages of the research.

4.7. Constructing the learner corpus

Ellis and Burkheizen (2005) suggest that implicit knowledge of language is most accessible in learners' spontaneous production, because it is in such conditions that their focus is on communicating the message and not on the language form. This assumption lay behind the next stage of the study: building and analyzing a language learner corpus (over 217,000 words). The material originates from blogs produced by 158 students over the period of four years as part of their IT course in their third year at KJO. Students wrote blog posts on topics of their choice or in response to tasks assigned by the teacher, in either case the aim being to share some experiences or reflections with other students, who could use the "comment" feature to react to what the authors of posts had to say. For technical reasons, these comments are not included in the corpus, but the interactive character of the project is undoubtedly reflected in the students' use of language, its style and level of correctness. The project's clearly stated principles included no corrective feedback on the part of the teacher so that students could feel free to apply the level of monitoring that they found adequate and sufficient. It must be added that the texts were produced outside the classroom, without any time pressure or supervision, so use of reference materials like dictionaries and grammar textbooks cannot be excluded. From the number and quality of errors recorded in the corpus one can gather that few students actually made intensive use of such reference materials, however. It has already been mentioned that the language which students used in their blog posts was in most cases rather informal and relaxed in style, so it seems that their use of monitoring was relatively low. With very few exceptions, however, the students refrained from using slang, offensive language or other objectionable forms.

The above observation might, in a way, confirm what Corder ([1976] 1981) claimed about language elicitation techniques and their influence on the data themselves. Generally, the data-collection method employed at this stage may be labeled as one of clinical elicitation, whose characteristic feature is that it is not controlled and does not limit the subjects' language production in any way. This technique allows the researcher to obtain more authentic data, which represent more closely the language that learners are likely to use in real communication. According to Corder ([1976] 1981: 69), "[l]earners typically produce a different set of errors in their spontaneously generated utterances, when attempting to communicate, than in their practice utterances". The

variation is supposed to replicate Widdowson's distinction between different types of rules that learners follow depending on how controlled their use of L2 is:

> I want now to make a distinction between two kinds of rule. The first kind, which I will refer to as *rules of usage*, account for linguistic competence in the Chomskyan sense: they represent the language user's knowledge of the formal systems of his language. We might say that they constitute his basic grammatical source of reference. The second, which I will refer to as *rules of use*, account for the language user's knowledge of speech acts and can be said to constitute his basic communicative source of reference (Widdowson 1979: 140).

What Corder ([1976] 1981) suggests is that in more spontaneous language production, learners follow *rules of use* rather than *rules of usage*, which affects the types of errors they commit. The distinction does not seem to be very clear-cut and the two sets of rules must both be active when learners speak or write in L2, but there indeed may be a difference in terms of which of the sets of rules dominates.

Finally, it is also worth pointing out that the corpus has been checked for quotations from literature, song lyrics and other types of 'external' texts, which were subsequently removed. In some cases short quotes essential to the syntactic integrity of the sentences in which they appeared were preserved.

4.8. Error analysis of learner corpus data

The corpus was searched for the same errors as those identified in the essays. The searches were performed with *TextSTAT* (Hüning 2008), a computer program which includes a concordancer, some basic statistical text-analysis tools, and a very useful export-to-MSWord capacity. The very nature of corpus studies, and hence corpus-based error analysis, is quite different from traditional error analysis. It requires the researcher to define a certain hypothesis, something s/he is looking for in the data rather than to browse the material sequentially in search of errors. This is what Granger (1998a: 15) calls "hypothesis-based research". The alternative approach, "hypothesis-finding research", where "the researcher may simply decide to gather data (…) and quantify everything he or she can think of just to see what emerges" (Scholfield 1995: 24, as quoted in 1998a: 15) is also an option with corpora. It is, however, much more successful with annotated corpora, where lemma, grammatical categories and patterns can be quantified

and not merely specific language tokens. In a simple text corpus, like the one compiled for the current study, such an approach needs to be limited to general corpus statistics, frequency lists and cluster analysis. For this reason, the hypothesis-based approach was adopted: the potential problem items were first identified by means of manual error analysis of the examination essays, and those problems subsequently determined the shape of the learner corpus queries. Some of the errors from the essays indeed reemerge in the corpus material, despite significant differences between the two collections of texts in terms of genre, style, and the level of control. Below is the list of some of the blog corpus searches for errors identified in students' essays. Only the base forms of words are listed, but searches were made for all forms of words by means of Regular Expressions Syntax (Friedl 2001), including expected incorrect forms. For example, to render the word *necessary* and its derivatives in all their predictable misspelled forms, the search term was formulated as follows: "*bnecc*ess**".[26] In some cases several searches were combined into one result. Not all of the errors were of the kind that simple unannotated corpus searches could render; many such errors had to be excluded from the project. In several cases, however, the results of PICLE searches were included instead. The texts in PICLE were written by learners of similar proficiency, though the level of formality of the language, the degree of supervisor's control over the situation, and hence the level of students' language monitoring were much higher in the case of PICLE. Still, some tendencies in students' choices of language forms may correlate, giving some insight into Polish advanced learners' problems with English. Some blog corpus searches revealed different problems from those identified in the essays. These have been noted as well, e.g. in the *articles* section, where apart from four missing article errors for *US/USA* (out of 66 hits), two agreement errors were found with this noun.

The table below is organized on the same principles as the error database. There are three major categories of error (form, lexico-grammar, and syntax), which are divided into the same sub-categories. It uses the following notions and abbreviations:

- hit – a single occurrence of a search term in the corpus (unless stated otherwise, the form given stands for the whole lemma, i.e. all forms of a given word);
- target – an erroneous/non-standard form which was the object of the search;

[26] Such notation allowed for various misspellings – a presence or absence of letters followed by the asterisk, and with the feature of the software "search whole words only" turned off, all words beginning with the root "necess" could be rendered. Span queries were also made with the use of RE, e.g "like(\W+\w+){0,5}\W+(do|does|did)" rendered *like* and forms of *do* within the span of five words.

- manual selection – in some cases there was a specific context required (e.g. part of speech or verb form); the blog corpus has no POS annotation, so some searches had to be filtered manually;
- OCs – obligatory contexts: contexts identified to be suited for the desired form; the tally includes the examples which contain errors;
- NN – PICLE annotation for a singular noun;
- VBZ – PICLE annotation for a singular form of *be*.

Table 11. Learner corpus error search results

Category	Sub-category	Form	No of hits	Target error	No of errors	Comments
FORM	categorial	mine	57	mine + noun	2	original error: **his or HERS photo*
		other	242	for *others*	4	used as a pronoun
				for *another*	5	used with a sing. countable noun
	graphophonemic	-aly	9	for *-ally*	5	1226 OCs
		beginning	68	begining	4	RegEx search: "begg?inn?ing"
				beggining	1	
		extend	9	for *extent*	4	essays: *prize* for *price* (lack of contrast as a result of final devoicing)
		side	47	site	2	
		intel(l)…	13	*intellectual/ intelligence, intelligent*, etc.	0	13 OCs
		…necessar…	24	…neccessar…,	2	RegEx search: "necc?ess?" (whole words search OFF)
				…necesar…	1	
		-nes	151	for *–ness* (*happines)	2	-ness – 148 hits (150 OCs)
		stuff	86	for *staff*	0	
		staff	4	for *stuff*	1	
		-uly	40	for *-ully*	1	76 OCs
	inflection	doors	4	for *door*	0	3 "funny English" quotes include the item
		of people	36	for *people's*	7	preferred form in the context: Saxon genitive
		todays	1	for *today's*	1	14 OCs
		others	72	for *others'*	3	4 OCs
		years time	11	for *years' time*	11	18 OCs (BNC: 217 non-standard/254 standard)
		's / s' (Saxon genitive)	568 /88	various misuse errors	26/ 7	wrong use of articles, wrong choice of form; preferred form: n+n or adj+n (corpus search + manual selection)
		x year(s)/ x month(s) old (+ n)	71	for: *x year/ x month old* (+ n)	3	phrase used nominally – no plural marking; corpus search + manual selection

FORM	style	single digits in text	739	in formal writing (essays): single numbers should be written in full	0	not identified as error - not stylistically inappropriate in blogs
	word formation	-ical	157	-ic	4	misapplied adjectival suffix (economic/economical, magic/magical)
		-ic (adj)	391	for *-cal* or Ø (e.g. tourist)	10	manual selection
LEXICO-GRAMMAR	agreement	information	31	plural verb or pronoun	2	3OCs (!) – the word seldom used as subject or referred to with a pronoun. Avoidance?
		majority	6	pl noun + sing verb	1	4 OCs
		most of	67	pl noun + sing verb	1	25 OCs
		N and N is/was	118	compound subject with singular verb	11	PICLE search: NN and VBZ [0,5]
		number of N	21	logical agreement problems	1	6 OCs
	article omission	human	29	article omission	1	6 OCs
				article misselection	2	
		speed (n)	6	article omission	0	3 OCs
		issue	16	article omission	0	12 OCs
		job	72	article omission	6	32 OCs
		level	24	article omission	6	11 OCs
		majority	6	article omission	3	6 OCs
		market	10	article omission	0	9 OCs
		mind (n)	102	article omission	4	7 OCs
		number	50	article omission	3	27 OCs
		opportunity	14	article omission	0	11 OCs
		problem	51	article omission	0	31 OCs
		rather + *adj n*	12	article omission	2	10 OCs
		side	47	article omission	4	35 OCs
		student	42	article omission	3	36 OCs
		such	193	article omission	13	86 OCs also: 3 misselections and 6 additions
		system	17	article omission	3	11 OCs
		teacher	236	article omission	11	193 OCs
		unknown	10	article omission	2	3 OCs
		US/USA (n)	59	article omission	4	only as head noun in NP; also: 2 agreement errors
		view (n)	28	article omission	1	13 OCs
		world	204	article omission	9	162 OCs
	colligation	allow for + n	33	for *allow + n*	5	6 OCs
		allow to + inf	3	no object	3	14 OCs
		approve	9	for *approve OF*	4	9 OCs
		chance(s) to + inf	28	for *chance(s) of + gerund*	2	2 OCs; avoidance? (no hits for *chance of* - BNC: of - 3126, to - 3905)
		decide on	2	used with gerund rather than a noun	0	2 OCs; (a different error: *decide OF +n*)
		doubt in existential structure	1	It is without doubt that…	1	1 OC (here: *There is beyond doubt…*)

LEXICO-GRAMMAR	colligation	difficulty|-ies	18	have difficulty|-ies to do sth	0	5 OCs
		example	47	example + wh clause (no prep)	1	4 OCs
		face (v)	19	**to face with sth*	0	14 OCs (face +n)
		force (0,4) to	24	force sb to + n **force children to education*	0	9 OCs
		say about	11	object omission	0	14 OCs
		such	14	such… like (for *such… as*)	2	RegEx search: such(\W+\w+){0,5}\W+ (as|like), manual selection
	collocation	acquire +n	4	acquire +n essays: *acquire *results*	3	errors here: *music*, *memories*
		contact (n)	13	*contact with* + inanimate n (e.g. **first contact with the computer*)	7	suggested alternatives: *encounter*, *experience*
		disappointment	8	v + disappointment: **stimulate disappointment*	0	1 OC
		example	29	adj + example: **negative example* (for *bad example*)	0	0 OCs
		law	18	v + law: **stand the law*	0	1 OC 2 word choice errors: *law* for *right*
		make	587	make + n (**make sth stupid*)	3	impossible to identify OCs – many options
		problem	144	v + problem: **discern a problem*	1	89 OCs
		right(s) (n)	15	v + right(s) **refuse the right*	0	15 OCs
		saying (n)	4	saying + v: * *a saying claims*	0	2 OCs
	tense/ aspect	if + would (sub clause)	5	*would* in past conditional	3	102 OCs (past conditional; manual selection)
		seem	6	*that* –clause (aspect)	0	0 OCs
			56	*to*-inf (aspect)	3	5 OCs
	word choice	ability|-ies	12	for *skill(s)*	4	20 OCs
		amount	12	for *number*	5	25 OCs
		can + v	162	for *may* +v	9	175 OCs
		careless	5	for *carefree*	3	10 OCs
		concentrated	4	for focused	3	7 OCs
		contemporary	6	style (too formal for the context)	3	suggested alternatives: modern, present-day, current
		during	136	for *in*, *at*, *on* in time expressions	36	a tendency to overuse *during* from L1 equivalent
		economical	1	for *economic*	1	7 OCs

LEXICO-GRAMMAR	word choice	idea(s)	150	v + idea(s) (**introduce* for *implement an idea* – covert error)	0	99 OCs (mostly *have, give, be*, but also *launch, present, promote*, etc.)
		It + 'be' (sent. initial)	282	reference error – preferred form: *This*	11	159 OCs
		This + 'be'	148	reference error – preferred form: *It*	7	289 OCs
		itself	28	for plural ref (*themselves*)	0	25 OCs
		join	26	for *combine*	1	9 OCs
		learn	155	for *teach* /	1	190 OCs
			155	for *study*	10	97 OCs
		possibility	13	for *opportunity*	5	17 OCs
				for *ability, power, being able to do sth*	2	15 OCs
		same + as (0, 3)	20	for *like* + n (taki sam/ten sam) **he is the same person as they are*	3	(*like*: 962 his) OCs not verified
		say	516	for *speak/tell/stand for*	1	214 OCs (tell)
					0	44 OCs (speak)
		some + sing n	50	used with sing count nouns	35	OCs impossible to establish – an optional/ emphatic feature
		state (v)	10	for *say, find, establish*	5	stylistic error – word much too formal for the context (indicates official character of the utterance)
SYNTAX	clause combining/ typographic	comma + *that*	131	comma splices with *that* as subject	17	There are 3161 hits for *that* in the corpus, most of which are used in noun clauses and relative clauses. Pronominal use seems least frequent.
				comma in noun clauses with *that*	14	
				def. relative clauses with *that*	4	
		like + do (0, 5)	31	comparison: **like he did* for *as he did*	5	24 OCs Note: the style in blogs is more informal so the form is not seen as an error here.
		after all	13	used as a conjunction – comma splice	1	6 OCs
		however	147	comma splice: to be replaced with a conjunction (*although* or *but*) or to be used as a linking adverb – with a semicolon/ full stop	17	OCs impossible to establish – many options
		send/sent	31	missing place adjunct (complex transitive str.)	0	17 COs
		prefer if	21	object omission: **I would prefer if…*	3	10 OCs
		worth	68		7	51 OCs

SYNTAX	parallel structure	and that	56	v +prep + n *and that*-clause *aware of.. and that...	5	15 OCs (*that… and that…*)
	voice	*-ing by* (0, 2) (faulty passive)	39	**achieving sth by someone* (preferred form: *sb achieving sth* OR *sth being achieved by sb*)	2	Polish calque: *osiągnięcie czegoś przez kogoś*
	word order	here	351	place adjunct between verb and object in verbs NOT referring to movement	28	56 OCs (*here* after an object or complement)
		hopefully	20	viewpoint adverb – awkward position, unclear modification: **…as soon as I graduate – hopefully.*	4	18 OCs
		not only	62	parallel structure (*not only.. but also…*)	6	36 OCS
				inversion	3	7 OCs
				postmodification, e.g. * *for the people of America and not only*	4	9 OCs suggested correction: *and others* OR (in other contexts) *and not only* ***that***
		wh-clauses		question WO in imbedded questions	30	OCs not calculated – too many options

The errors which were found to have a relatively high frequency of occurrence in the blog corpus, or to be interesting for other reasons, are discussed below. Several concordances from the corpus are included in Appendix B.

4.8.1. Errors of form

Grapho-phonemic errors: Obtaining data for misspelled words in a corpus cannot be done through a systematic search. However, as mentioned above, the most prominent grapho-phonemic errors in the essays were problems with double-letter spelling, and the adverbial suffix {-ly} was associated with the biggest number of such problems. The mechanisms involved in the occurrence of these errors have been accounted for in section 4.6.1.1. The same errors can be seen in the corpus as well, together with some quantitative data: out of 3608 words ending in {-ly}, 3042 were either adverbs or adjectives (*only* was excluded as a highly frequent but risk-free word), and 80 were spelled

incorrectly. This is not many (2.6%), but the errors do shed some light on students' spelling problems. Some '*l*'s are not doubled although they should be (**hopefuly*, **mentaly*, **vitaly*), others are, although they should not be (**fortunatelly*). Sometimes the doubling is applied to a different consonant (**finnaly*). The same set demonstrates problems with the spelling of reduced vowels, observed in the essays too: **basicly*, **chronicly*, **completly*). The three dominant errors are **definately*, **(un)fortunatelly* and various erroneous spellings of *necessarily*. Perhaps explicit class instruction on how these words should be spelled, supplemented with intensive practice, could save students and teachers a lot of frustration here.

Errors like *prize/price*, which appeared in the essays, were also found in the corpus, but with a slight variation: in four cases out of seven, the noun *extent* was spelled like the verb *extend* (see Appendix B). The source of the error may be the same as in *prize/price* (lack of a contrast between word-final voiced and voiceless obstruents in Polish), except that here it operates in the reverse: a letter for a voiceless obstruent /t/ is replaced with one for its voiced counterpart /d/.

Inflection errors identified in the essays most often involved the use of the Saxon genitive. The blog corpus material contained 568 instances of the Saxon genitive in its base form ('s), 26 of which contained errors (4.6%) with 88 instances of the Saxon genitive marked only with an apostrophe, to be used with regular plural nouns, 7 of which were erroneous (8%). The errors involved various grammar problems: the choice of Saxon genitive where the preferred form would be a compound noun phrase (**camera's flashes*, **a computer's keyboard*, **math's teacher*); the use of a base form Saxon genitive with nouns which should be plural in form (missing plural marker: **the Polish Teacher's Union*); problems which result from a wrong NP structure or overuse of the definite article (e.g. **my auntie's Beata place*, **the Carroll's novels*, **the Obama's presidency*). Finally, two errors seem purely typographic and are known to appear in native speaker English as well (18 hits in the *BNC*): the use of an apostrophe after the possessive determiner (**its'*).

As the above analysis demonstrates, a more general problem emerges with using the Saxon genitive for plural nouns. In these cases omission seems more frequent than elsewhere. To confirm this, a POS search should be carried out, which could identify two nouns in a sequence, the first plural and the second either singular or plural. Such a

search, however, can only be carried out on a POS annotated corpus, a feature absent from the blog corpus. Instead, a search was performed on PICLE, and some interesting observations were made. The search syntax was defined as "NNS NN*" (a plural noun followed by a singular or plural noun) and yielded 318 results, among which there were 31 errors of the Saxon genitive omission. Other searches identified 355 correctly structured instances of the plural Saxon genitive. Altogether then, there were 386 (355+31) obligatory contexts for the plural Saxon genitive, in 31 of which (7.9%) errors appeared. Considering that the texts were produced by advanced learners, this seems a high error statistic for a relatively simple language feature. It may be that the origin of these errors is intralingual: at a certain point learners find out that the Saxon genitive is used in a reduced form with plural nouns, and then they misapply this rule. The fact that this form of marking inflection is absent from Polish may still contribute to the high frequency of these errors.

Another comment is due with reference to the temporal use of the Saxon genitive in phrases like *in five years' time*. Although the apostrophe is the norm here, native speaker English demonstrates a decline in its use: the *BNC* records 254 uses of *years' time* against 217 of *years time* (no apostrophe). Considering that the corpus texts under scrutiny originate from blogs, which generally assume a lower level of formality, the high frequency of the form without the possessive marker (11 out of 18 instances *of years(') time*) is not surprising and will not be seen as an error here.

Word formation: The most prominent errors of word formation observed in the corpus are connected with the suffixes: {-ic} (for adjectives). There were 391 adjectives ending in {-ic}, out of which 14 were incorrect (3.6%). Most of the errors may be traced to the Polish derivative suffix {-czny} (**touristic*, **architectonic*, **logistic*). In some cases the errors could be attributed to an underdifferentiation of two English words which have the same Polish equivalents, e.g. *magic* ("able to make impossible things happen") vs. *magical* ("full of mysterious quality" – both definitions Macmillan 2007), both of which have one equivalent in Polish: *magiczny*.

4.8.2. Lexico-grammatical errors

Agreement: The errors found in the essays were not very prominently present in the corpus. Single instances of agreement errors with *majority of, number of* and *information* were found. This does not mean that agreement errors do not appear in the corpus, but if they do, they must involve different lexical items or structures that cannot be retrieved through an unannotated corpus search.

Article omission: Corpus searches for errors of article omission were carried out for nouns which proved problematic in this respect in the essays. Obligatory context for the use of *any* article is understood to be the use of a singular countable common noun without any other determiner. Some of the results are given below:

- level – 24 hits, 11 obligatory contexts, 6 article omission errors (54.5%);
- majority – 6 hits, 6 obligatory contexts, 3 article omission errors (50%);
- job – 72 hits, 32 obligatory contexts, 6 article omission errors (18.8%);
- teacher – 236 hits, 193 obligatory contexts, 11 missing-article errors (5.7%);
- such (a) + noun – 193 hits, 86 obligatory contexts, 13 article omission errors (15.1%). Also: 6 article addition errors – indefinite articles added in front of uncountable nouns, probably an induced error, resulting from strong focus on the pattern of "*such* + *a* + noun" in EFL grammar instruction.

Colligation: The search for errors of colligation has produced quite interesting results. First of all, many of the errors identified in the essays were not present at all in the corpus. The verb *face* has been used quite often and without problems in complementation, which did appear in the examination essays where students used *with + NP* instead of an NP alone. On the other hand, its more formal synonym *tackle*, which was a source of quite a few errors in the essays, appears only once and is used correctly. This indicates the use of the avoidance strategy – choosing the word the learner feels safer with, especially in a context which does not require a formal style. Avoidance does not account, however, for low error rates in all the corpus observations. As far as the use of *have difficulty/-ties (in) doing sth* is concerned, neither avoidance nor erroneous usage has been observed (18 hits, no error). The only problem here is the overuse of the plural

form, compared with proportions of *[have] ... difficulty / [have] ... difficulties* (span: 3) in the *BNC*:

Table 12. Corpus statistics for *[have] difficulty/difficulties* (span: 3)

	BNC	blog corpus	PICLE
[have] ... difficulty	1.65/100k[27]	1.38/100k	0.30/100k
[have] ... difficulties	0.42/100k	1.38/100k	3.89/100k

PICLE statistics are even more strongly skewed in this way. The imbalance may result from interference from Polish, where the equivalent phrase uses the plural noun (*mieć trudności*).

Similar problems emerge when students' use of the word *chance* is analyzed. The word appears 43 times in the corpus, out of which 27 times it is used with an infinitive (only once incorrectly), and there is not a single use of the word with *of + noun/gerund* complementation in the data. This is in strong contrast to *BNC* data again, where the two forms are almost perfectly balanced:

Table 13. Corpus statistics for *chance of/chance to*

	BNC	blog corpus	PICLE
chance to + infinitive	3.78/100k	12.43/100k	11.96/100k
chance of	3.13/100k	0.00/100k	6.87/100k

The high statistic of *chance to* in blogs might have been considered insignificant on the grounds that blog writing has unique stylistic features, but PICLE data seem to confirm Polish students' tendency to overuse this form, although the imbalance is not so severe there. The total absence of *chance of* in the blog corpus is, however, very surprising.

Another word identified as a source of errors in the corpus is *allow*. Lewandowska-Tomaszczyk explains Polish learners' problems with using the word as follows:

> One of the reasons is that in a number of contexts where different senses of *allow* and *permit* are used in English, in Polish, the verb + preposition construction *pozwolić na coś (komuś)* can be used, even though the sense would not be always equivalent to the English one. ... Even quite advanced students of English commit errors identifying the English *allow for* 'take account of' with the Polish *pozwolić (na coś)* lit. 'allow for' in the sense of 'permit'. The source of the error is the absence of such semantic contrast in Polish (Lewandowska-Tomaszczyk 2001: 215).

[27]The statistics are given in number of items per 100,000 words in the corpus.

As the corpus search results demonstrate (see Appendix B, section 1.5), out of 33 uses of *allow* in the corpus, five (15.2%) include exactly the type of error discussed above. In another three there was an object missing in front of the infinitive (e.g. **some software that allowed Ø to mix together sound samples*).

The last case worth discussing in this group of errors is the use of *approve*. There are nine (9) instances of the word in the corpus, and in four (4) of them the preposition *of*, which should be used there, has been omitted. The structure *approve +NP* is grammatical in English (e.g. *to approve a plan*), but does not have the meaning that the students intended to express (*approve of something* – "to think that someone or something is good, right, or suitable" – *LDOCE5*). On the surface, the problem seems to be one of complementation and as such it was classified, but considering the kind of confusion that the two options cause, it borders very closely on word choice. Since the verb with and without the preposition have such distinctly different meanings, they should be seen as two separate lexical items. This is another instance in which the borderline between lexis and grammar is very blurred and indiscernible.

Collocation: Surprisingly perhaps, errors involving collocation do not feature very strongly in the corpus data; at least the ones which occur in the essays are not very frequent in the blogs. There are a few errors, however, which have indeed been found there. First the use of the noun *contact*: There are 13 instances of its use in the corpus, out of which seven (7) are problematic (53.8%). Most of them result from the fact that one of the Polish uses of the phrase *kontakt z czymś* (lit. 'contact with something') refers to an experience of using something. English does not use *to have contact with* in the same way. The prepositional complement needs to have personal, personified or abstract reference, in which case the meaning of *contact* is close to *communication* or *connection*. Alternatively, the noun that follows can refer to a substance or pathogenic organisms, in which case a notion of physical contact is involved. Hence sentences like these from the *BNC*:

Everyone will have had contact with the police, if only to ask directions.

Fieldwork brings an indispensable first-hand contact with problems.

The chemical was liable to explode on contact with water.

The errors identified in the corpus usually involved references to the computer, because one of the assigned topics was "My first computer experience". For example:

**My first contact with the computer was when I was about ten.*

No instances of such or similar uses were found in the *BNC*, other credible online corpora or in dictionary examples.

The corpus errors of collocation with *contact* as a noun are, therefore, different from the one error observed in the essays (*make contact* used instead of *develop relations*), but their quantity indicates that the word needs more attention in the classroom.

Finally, students proved to have some problems with using another common word in phrases, i.e. the word *time*. In the essays one of the errors was **in the right time*. Search in the blog corpus rendered 915 uses of *time*, and a complex search of *in* occurring with *time* in the span of two words in between (IN + TIME – 0, 2) revealed five similar errors – involving a choice of the appropriate preposition with the word *time*. The problem is that in each case a different preposition was required (see Appendix B, p. 379). The Polish preposition in these contexts would not always be *w*, the equivalent of *in*, either. Therefore the errors do not seem to have any L1 sources, and so are not included in the table above.

The form **enroll to* does not have such a strong presence in the blogs as it does in the essays: the lemma *enroll* was used only three times in the corpus, and in one of these cases the incorrect preposition was used. The sample is so small that any generalizations as to the error's frequency in the population would have to be very tentative here.

Word choice errors: A lot of errors of word choice identified in the essays found their confirmation in the corpus. The most prominent ones are discussed below:

- *careless* – this word or its variants (*carelessly*/*carelessness*) appears five times in the corpus, and three of them (60%) are used with positive meaning, which is exactly the same error as those in the essays. The distinction between *careless* and *carefree* definitely needs teachers' attention, as the error may cause confusion and communication problems.
- *possibility* – this noun is notorious among Polish teachers of English as one of those which causes numerous problems for their students. This is also the case in the corpus data. Out of 13 occurrences of the word, six are flawed in terms of use (and one in terms of spelling). Problems with using the word originate from the Polish *możliwość*, one of its possible translation equivalents, which has a much broader

meaning than the English *possibility* (where it is mainly associated with likelihood or potential). *Możliwość* includes such meanings as *opportunity*, *ability* and *power to do something*. This causes a lot of usage problems and makes it difficult for students to understand the difference, and, more importantly, to make these distinctions in their language production. In the data three instances of *possibility* being used should actually have had *opportunity* instead, and the other three would have been easily corrected by replacing *possibility* with *ability*. Potentially, however, there are many other ways of translating the word *możliwość* into English, and students need to be made aware of this. Another problem is the complementation of the word. According to *LDOCE5*, the form required here is *of/for+NP* or, far less frequently, *that* + clause, while Polish students tend rather to use the infinitive. This may be a less obvious L1 effect, as according to the search results in NKJP (the Polish National Corpus)[28] the word *możliwość* is most commonly followed by a gerund and not an infinitive. Still, the infinitive is a possible choice in purpose constructions, so interference can be considered a possible explanation, especially in that the error does not appear in a Hungarian learner corpus of English, for example. Kaszubski (1999) also indicates Polish influence as the source of the error. Admittedly, similar problems are reported to appear in German learners' English (Mair 2002: 121), except that the German equivalent of *possibility – die Möglichkeit* – has a similar range of meanings to the Polish *możliwość* (*possibility*, *opportunity*, *chance*, and, in plural, *capabilities*), according to *Collins German English Dictionary* (2011). Judging from the examples provided[29], *die Möglichkeit* can also be followed by an infinitive: "die Möglichkeit haben, etw zu tun", which may indicate that German learners have very similar interference problems with the word as Polish learners do.

- *some* + singular countable noun – The use of *some* in front of a singular countable noun is not always an error. The form does exist, but learners use it much too frequently as a marker of indefiniteness instead of the indefinite article, which is the default form associated with this function in English. Native speakers of English use *some* as a marked, emphatic form, stressing the indefiniteness and usually carrying negative prosody. Polish learners tend to translate the Polish lexical marker of indefiniteness *jakiś* into *some*, probably for the same reason for which they omit indefi-

[28] Collocation search url: http://nkjp.uni.lodz.pl/?q=2wger2z
[29] The entry is available online at http://www.collinsdictionary.com/dictionary/german-english/möglichkeit

nite articles in front of singular countable nouns – they often do not seem to understand or are not able to apply the concept of the indefinite article. Blog corpus data give 50 instances of various singular countable nouns with *some*, 37 of which do not seem appropriate (74%). Those which do are part of phrases (e.g. *some day*, *to some degree*, *on some level*) or have a justified emphatic function (e.g. *show me some proof of...*, *some unexplained twist of hormones*, *some guy*). Many, however, do not seem to have any justification (e.g. *sitting in some open window*, *I grabbed some book by...*, *qualify for some championship*). The same problem features in PICLE, although the statistics there are slightly less striking: *some* appears in front of countable nouns 79 times, 27 of which (34.2%) contain inadequate uses of the determiner. It is also worth comparing learner corpus search results with those from a native speaker corpus. Table 14 below presents a comparison of *BNC* and blog corpus statistics for a few such uses and their equivalents with the indefinite article:

Table 14. *BNC* and blog corpus statistics for indefinite singular countable nouns

	BNC frequency	blog corpus frequency	*BNC* per 100k	blog corpus per 100k	blogs/*BNC* proportions
a book by	321	4	0.33	1.84	5.58
some book by	0	1	0.0	0.46	-
a championship	59	0	0.06	0.0	0.0
some championship	0	2	0.0	0.92	-
to a degree	209	0	0.213	0.0	0.0
to some degree	252	2	0.26	0.92	3.54
a * school	1534	21	1.56	9.67	6.20
some * school	17	1	0.017	0.46	27.06
	(1)		(0.001)		(46)
a task	517	4	0.526	1.84	3.50
some task	13	1	0.013	0.46	35.38

As it was mentioned in the previous chapter, comparisons between corpora that are very different in composition and size need to be considered with caution, but the differences in the proportions between *a/an* and *some* in the two corpora may reflect a trend in Polish learners' choices. The numbers that are probably worth most attention are presented in the last three columns. They show the proportions between the blog corpus and the *BNC* in terms of how frequently *some* and *a* are used in front of singular nouns. For example, the sequence '*a * school*', i.e. 'indefinite article + any word + *school*', appears 6.2 times more often in blogs than in the *BNC*, but the structure '*some * school*' – 27.06 times more often, if the sizes of the corpora are

considered in the calculation. What is more, in the few examples that do occur in the BNC, *school* is merely a modifier of a plural or uncountable noun that follows; only in *one* example is *school* used as the head noun (hence the bracketed number in the table above). The higher representation of some phrases in the blog corpus results from the dominance of educational topics covered there and the fact that the *BNC*, as a balanced corpus, does not contain such a large body of education-oriented texts. These disproportions are minimal, however, if compared with those of *some* vs. *a* in front of singular nouns. They are clearly the highest where the *BNC* does not have instances of the former and the blog corpus does, but for these the quotient cannot be calculated (division by zero). A full list of the blog corpus concordances of *some* in front of countable nouns can be seen in Appendix B, page 382, with instances judged as being inappropriate underlined.

- *amount/number* – Although the error of distinguishing between these two words does not appear in the essays, a search for instances of errors of article omission in front of *amount* gave evidence of this being quite a common problem. Students tend to choose *amount of* with countable plural nouns, instead of *number of*, a standard choice in a context like this. It seems that the problem originates from the lack of such a clear-cut distinction in Polish, which does have two corresponding words *ilość* ('amount') and *liczba* ('number'), but the former seems much more universal as it can be used with both countable and uncountable concepts.[30]
- *It/This* distinction – The reference system in English causes numerous problems for Polish learners of English. Use of articles is the most notorious one, but there are others. Some errors involve using inappropriate pronouns for anaphoric reference. The rules presented below, concerning *this*, are not always followed by learners:

> Words such as *this*, *which* and *such* can also be used to refer endophorically to segments of the text longer than just a noun phrase. That is, they may refer to preceding or following words and phrases, a previous or upcoming whole sentence or, sometimes, a whole stretch of text:
>
> > *He is very experienced.* This *is why we should invite him.*
> > (Carter and McCarthy 2006: 245)

[30] This can be confirmed by going to a sample search in NKJP available at http://nkjp.uni.lodz.pl/?q=26ykdc4

In the sentence above, even quite advanced Polish learners would often use *it* instead of *this*. Differences between *it* and *this/that* are further defined as follows:

> The impersonal pronoun *it* and the demonstrative pronouns *this* and *that* are used in different ways to refer to segments of text or to ideas in the text. *It* is used to continue reference to an entity which has already been established as a topic in the text:
>
> *The girl was so ecstatic afterwards, she had such a wonderful smile on her face. It was a sight I shall never forget.* (*It* refers neutrally to the girl's ecstatic behaviour and smile) (Carter and McCarthy 2006: 246).

In other words, there are two main options. One is reference to a noun phrase introduced earlier in the text and already brought into focus, in which case the reference pronoun is *it*. The other is reference to a fragment of text larger than a noun phrase (an idea expressed in a whole clause or sentence, for example) and intended to become a new topic of the discourse, where the preferred reference item would be *this* (or – less commonly – *that*). It must be added that once the new topic is established, reference to it is continued with the impersonal pronoun *it*. Corpus search for such errors was limited to sentence initial use of *It* followed by *is, was* or *has.* The search produced 282 results, out of which 12 proved problematic. Each of the cases (presented in Appendix B, p. 383) has been analyzed in a broader context, so that the kind of reference involved could be established, because the KWIC format does not usually allow such judgments to be made due to the scarcity of context. For example, an analysis of a broader context for the first line in the concordance shows that it indeed contains a reference error:

*Namely, she's learned to enjoy life while it lasts because come June it could be over. *It has made her stronger and more independent, since all throughout this time she was more or less alone.*

The use of the reference item *It* indicates the noun *life* to be its antecedent, whereas in fact the reference is made to the message of the whole previous sentence (the fact that she learned to enjoy life), so *This* would be a much better choice.

- *learn/study/teach* – This group of errors is especially important considering the fact that students participating in the project were studying to become teachers, so educational issues involving frequent use of the words in question were in focus all the time. The error can be accounted for by underdifferentiation: the three English

words have one Polish equivalent – *uczyć*, the first two of them in the reflexive form (*uczyć się*). The distinction between the two is very subtle, but quite important in educational contexts. In *learn* the focus is on the final outcome (gaining some knowledge or skill), whereas in *study* the emphasis is on the learning process, the activity itself. In the essays there were three cases of *learn* being used instead of *study*. The blog corpus provides more information about the frequency of such errors: the lemma *learn* occurs 155 times there, out of which 13 should have been replaced with *study* (9%) and one with *teach*. This seems an incredibly high percentage considering the circumstances.

- *to state* – This verb is commonly overused by Polish learners as it bears a superficial graphophonemic similarity to the Polish verb *stwierdzić*, which is used with a wide range of meanings and is often equivalent to such English verbs as *find* (especially in academic context), *claim*, *establish*, *say* and, indeed, *state*. Students do not recognize the uniquely formal character of the English verb, which is confirmed by corpus data: there are 50 hits of the lemma *state*, 10 of which are verbs, and among these four are used in a stylistically awkward manner.
- *during* – This preposition of time is oddly popular with Polish learners, who tend to use it by means of a calque in phrases where Polish uses *podczas.* The problem is that in English time expressions the choice of a preposition often depends on a particular noun to which it is bound, so the concept of duration, implied in the word *during*, is not enough for it to be actually used. Hence forms like *during Christmas*, far less frequent in the *BNC* than *at Christmas*, or *during classes* rather than *in class*. Generally, the standardized frequency of *during* in the BNC is 43.53 per 100k, and in the blog corpus – 63.09 per 100k, i.e. much higher. Since *during* is a function word, its frequency does not seem to be very dependent on style and genre, and such a big disproportion may indicate its significant overuse by Polish learners.
- *economic/economical/economy/economics* – It was somewhat surprising to see that only one out of fifteen occurrences of these words in the blog corpus was erroneous, because these errors are known to be very common in Polish learners of English, and they also appeared in the examination essays, even though the body of text was much smaller there than in the blog corpus. In the essays the confused words were *economical* and *economic*, but here the error was slightly different: **Economical University* instead of *University of Economics*. It may be that there was so much

emphasis on the former distinction in class that students eventually ceased committing the error, especially if not under stress.

4.8.3. Syntactic errors

Identifying syntactic errors in the corpus would have been much easier and more effective if the corpus were annotated, but even in its simple text format some interesting observations were made:

Clause combining – Searches were limited to particular conjunctions or words used mistakenly as conjunctions in the essays:

- *like + do* – As discussed earlier, a common error in comparative clauses involves a use of the preposition *like* instead of the conjunction *as*. There were five (5) instances of such usage (the search was limited to comparative clauses with the auxiliary *do*, span 5), compared with 22 cases of a correct choice of connector (*as +do*). The problem with this error is that it frequently occurs in native speaker English, usually in informal contexts (the BNC search gives 514 results for personal pronoun subjects only). Its occurrence in learner English may therefore result either from TL input or from L1 interference, or both. What is important is for learners to realize that the form is not appropriate in formal contexts.
- *who/which* – Only 2 out of 495 instances of *which* were used incorrectly instead of *who* with the intention of human reference.
- *however* – Problems with using *however*, identified in the essays, have found their strong confirmation in the corpus data. There are 147 hits for *however*, out of which 92 cases are capitalized and the other 55 are uncapitalized. The capitalized tokens are excluded from considerations of comma splice problems as they are sentence initial. A closer analysis of the uncapitalized cases revealed 17 comma splice errors (11.6% of the total corpus frequency).
- comma + *that* – As in the essays, the inappropriate use of a comma in front of *that* was very common in the corpus. There were 131 cases of *comma + that* in the data, out of which 37 were incorrect (28.2%). Those judged as acceptable involved embedding or apposition. What was surprising was that 17 out of the 37 errors involved

instances of a comma splice (again), where *that* as a demonstrative pronoun is the subject of the clause that follows the misapplied comma. Noun clauses, which were expected to be the dominant structure in this category, appeared in 14 cases, and relative clauses in four. Two examples include an unnecessary comma in *now that* as a time conjunction.

Clause structure errors – As previously discussed, errors of clause structure usually involve the absence of obligatory sentence elements like subjects, objects or adverbials (with complex transitive verbs) or faulty structures associated with particular sentence functions (conditional, reason, purpose, comparison, etc.). It must be said that these are not easily retrievable in a corpus search.

- subordinate subject omission – This error is difficult to diagnose in an unannotated corpus, but searches for the subordinating conjunctions which were used in the essays near such errors showed that their number in the corpus was minimal:
    - *if* – 513 hits, 1 subject omission error
    - *because* – 462 hits, 1 subject omission error
    - *how ... is/was/has* – 59 hits, no subject omission errors
      regular expressions search: how(\W+\w+){0,3}\W+(is|was|has)
- object omission – Essay analysis showed one object omission error, with *prefer*. This was not confirmed as a potential problem area in the corpus data, which does not exclude the possibility that the same problem could appear with other verbs.

Active/passive voice – An error which often goes unrecognized by Polish teachers of English is the use of *gerund +by+NP*, where the NP refers to the agent of the action expressed in the gerund. In the essay material the problem was diagnosed in the following sentence: **Educating by parents was very popular...* This is an attempt at building a passive construction with a gerund, but it would require reconstructing to become acceptable in English: *Children being educated by parents* or *Parents educating their children*, although in this case a deeper modification would be necessary for the utterance to sound more natural in English (*home schooling*). The error probably arises through a calque, because there is nothing ungrammatical about structures parallel to this in Polish, as in the following fragment from the NKJP: *przyjmowanie przez nauczycieli drogich prezentów.* A search in the blog corpus (search syntax with regular

expressions: “ing(\W+\w+){0,2}\W+by\W”) yielded 39 hits, of which two contained the erroneous structure in question. This statistic does not indicate that the problem is particularly urgent, but many more such errors occur when students are assigned Polish sentences with such phrases for translation.

Word order errors – a corpus is a useful tool to diagnose word order errors connected with the use of particular lexical items, of which there are quite a few in English.

- *here* – Errors of word order with *here* are quite frequent in Polish learners’ English, which is reflected in the blog corpus data. There are 351 hits for here in the blog corpus (for non-capitalized spelling), out of which 28 have been diagnosed as faulty in terms of word order (8%). Most of the errors occur with locative complex transitive verbs, which should be followed by an object and then an adverbial of place (e.g. put something somewhere). In the sentences identified as incorrect *here* immediately followed the verb, and the object came later.
- *not only* – Students have various syntactic problems in using the phrase *not only*. The ones identified in the essay material were mainly connected with inversion, but the corpus reveals other issues, too. When *not only* appears sentence-initially, problems with subject-verb inversion appear again. If it is used inside the sentence, there are problems with keeping the structures parallel. Also, there are a few cases of the phrase appearing at the very end of the sentence, which sounds awkward and of which there is no evidence in the BNC. Generally, students should be made aware of all these difficulties and trained appropriately in using the phrase correctly.
- *wh*-words – Many word order errors in the corpus involve the use of question words. This is a result of a feature characteristic of embedded questions and other *wh*-clauses in English, which do not have auxiliary inversion as direct questions do. Such complex distinctions are absent from Polish, where question formation itself does not have a syntactic marker other than a question word attached sentence-initially (and even that is sometimes optional in yes/no questions). This causes a lot of difficulty for Polish learners and is reflected in the 32 errors identified in the corpus by means of a complex search for non-capitalized *wh*-words (*when*, *where*, *what*, *why*, *how*), which rendered 2017 results. A percentage cannot be calculated on the basis of these frequency data, because not all results constituted an obligatory context for the rule in question to be applied. The rule is not always strictly followed

by native speakers of English as well, especially in speech, but in formal writing students should be cautious about this problem.

4.9. The experiment – the first stage (2009/2010)

Having established problem areas characteristic of the target group, the author undertook the experimental phase of the research (stage 4 on page 199) in the academic year 2009/2010. It turned out, however, that the two groups of subjects engaged in the project differed significantly, which was not addressed in the design of the study. That is why the first stage did not bring satisfying results and the procedure had to be repeated the following year, after necessary modifications were introduced. Although the first experiment was not entirely successful, it did offer some insight into the subject matter of the study, and so is presented briefly in this section as a 'first stage'.

The design was as follows: two groups of students (A and B) in the third year grammar course were taught lessons on the same language problems in two different ways – corpus-based and without corpus materials. Since, for reasons already presented, the groups were very small, a whole series of lessons was planned rather than a single experiment, so as to increase the validity of the results. To make sure that possible inequalities between the groups did not affect the outcome of the research, the following research plan was employed:

Table 15. Research plan

	experimental group (corpus)	control group (no corpus)
language problem 1	A	B
language problem 2	B	A
language problem 3	A	B
language problem 4	B	A
language problem 5	A	B
language problem 6	B	A

The effect of the treatment (data-driven learning techniques) was measured for each experimental group separately, and compared with the results obtained by the other group, which worked on the same language problems using other, more conventional, techniques. The assumption was that if data-driven teaching is more effective than its

alternatives, the test results for particular language items would be better (or to be more precise the number of the targeted errors would be smaller) in the group in which those items were taught with the corpus based materials. Language problems were grouped into pairs similar in subject matter and level of difficulty, so that each group would have a similar learning experience after the whole sequence of lessons.

Already at the initial stages of the experiment some problems emerged: there were noticeable differences in the two groups' level of English as well as their degree of willingness to work in less conventional ways. Their practical English examination results were compared and it became evident that one of the groups was much stronger than the other. To minimize these differences and achieve some balance, the researcher had to exclude some students from the experiment, according to recommendations made by Dörnyei (2007). Students with the highest results in the 'better' group and the lowest results in the 'weaker' group were eliminated from the study. The decision was based on the final practical English exam results, which was a broad battery of tests that offered a detailed diagnosis of the students' language skills as well as their command of grammar and vocabulary. Their exam averages were 66.9% for group A and 68.2% - group B. As a result of this selection, the original gap in the test means between the two groups could be reduced from 4.3 percentage points to 1.3 percentage points, which a *t*-test showed to have no statistical significance ($p = .24$), indicating that the groups should be comparable. Needless to say, the excluded students still attended the classes, but their results were not included in the study. Thus the established groups, each of which was composed of 11 students, were no longer significantly different in terms of their proficiency in English (as represented by their exam results). However, other differences – mainly those of attitude – remained, and it was these that could well have affected the results of the study.

Six language problems were presented to the two groups in this manner, which means twelve different lessons were carried out; each group had three corpus-based classes and three conventional, non-corpus lessons. There were three tests carried out, which checked how well particular language problems included in the experimental lessons were learned by the students. Each language problem was tested by five items in a test. Calculations were made to obtain a mean error rate in each group for items taught in the corpus-based and in the control lessons. The means were analyzed in a t-test, to

make sure the differences between them were statistically significant, which indeed was the case with *p*-values for the more conservative two-tailed tests well within α of 0.05.

Table 16. Two-Sample t-test first stage: A-corpus, B-non-corpus

	Gr A - corpus	Gr B - non-corpus
Mean	1.36	0.70
Variance	1.30	0.84
t Stat	2.62	
Significance (two-tailed)	0.011	

Table 17. Two-Sample t test first stage attempt: A-non-corpus, B-corpus

	Gr A – non-corpus	Gr B - corpus
Mean	1.91	0.97
Variance	0.77	0.97
t Stat	4.09	
Significance (two-tailed)	0.00012	

As the data demonstrate, the results are inconclusive, to say the least. Regardless of the procedure employed in the classroom, group A proved to have made almost twice as many errors in the tests than the other group. It transpired that the differences between the groups, which seemed to have been neutralized by selection criteria, were deeper than the level of English evidenced by the examination results. Any possible advantage one teaching technique could have had over the other was outweighed by the discrepancies between the groups at some other levels which were not controlled in the study.

Controlling learner-related variables is often a problem in second language acquisition research, as admitted by Jarvis (2000). In his research into L1 influence he managed to control eight out of a list of nine variables that could be involved in such a study, and "[t]he only factor not held constant or actively investigated through a balanced design is the second category: personality, motivation, and language aptitude" (Jarvis 2000: 261). The problem is that these learner characteristics, which are very difficult to measure and control, may have an effect on the outcome of an individual's effort at language learning.

Some light on the results was cast by responses to a survey which was supposed to collect information about students' impressions of and attitudes towards data-driven learning experience. While the two groups did not differ significantly in their evaluation of the effectiveness of corpus-based techniques, their attitudes were significantly different. The question asked was as follows:

> Choose ONE of the statements below that best describes your attitude to using concordances in class. Indicate your choice by circling the letter that precedes it.
>
> A. I didn't like working with concordances initially, but later I changed my mind.
> B. I liked working with concordances initially, but later I changed my mind.
> C. I didn't like working with concordances from the beginning and haven't changed my mind.
> D. I liked working with concordances from the beginning and haven't changed my mind.
> E. None of the above (explain briefly)
> .....

The choices were assigned Likert scale values as follows: C – 1, B – 2, A – 3, and D – 4. Open answers in E were adequately converted to this scale depending on how positive or negative they were. The two groups' averages of responses differed significantly (*p*-value: 0.026):

- group A – 2.27
- group B – 3.36

The diagram below shows how individual students' answers differed between the two groups. Points closer to the middle of the diagram indicate negative attitudes.

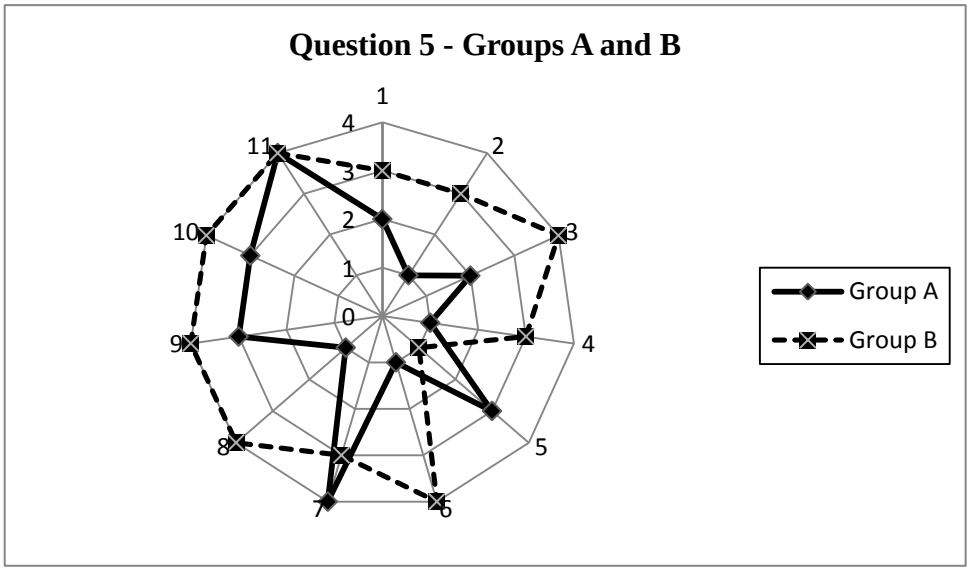


Figure 17. First attempt - survey question 5

Another of the survey questions was: "*Tick YES or NO next to the words/phrases listed below to describe how you assess corpus-based activities.*" A list of eight adjectival expressions followed (see table 8). A "*yes*" was counted as one, a "*no*" – as zero; hence all the results are given in decimal fractions on the scale from 0 to 1. There was a third option: "*I don't know*", so that students would not be forced into opinions they did not hold. The numbers in the "N" columns in Table 18 refer to the number of students who gave a full-value answer to this question, rather than "*I don't know*". The gross averages for all positive vs. all negative adjectives allow some insight into the groups' attitudes, especially if differences between those averages are considered. Group B's mean for negative adjectives is lower than group A's, but for positive adjectives – much higher. As a result, the gap between the two scores for group B is nearly twice as big as for group A (0.66 and 0.34 respectively), which proves a much more positive and open attitude among its members. The biggest differences between the groups are observed in the case of responses to three adjectives: *convincing* (gr A: 0.56 vs. gr B: 0.80), *effective* (gr A: 0.75 vs. gr B: 1.00) and *difficult* (gr A: 0.44 vs. gr B: 0.09); again, in each of these cases group B proves to be more positively disposed towards data-driven learning techniques.

Table 18. First stage - survey question 3

ADJECTIVES	N Group A	N Group B	Mean Group A	Mean Group B
convincing (+)	9	8	0.56	0.80
thought-provoking (+)	9	9	0.78	0.89
illuminating (+)	9	9	0.78	0.90
effective (+)	8	9	0.75	1.00
POSITIVE – mean value			0.72	0.90
overwhelming (–)	11	9	0.36	0.20
difficult (–)	9	10	0.44	0.09
time consuming (–)	11	7	0.27	0.29
boring (–)	11	10	0.45	0.36
NEGATIVE – mean value			0.38	0.24
Difference (positive-negative)			0.34	0.66

There is a strong indication that these differences in attitude and the discrepancies between the error statistics are connected. The one question which is difficult to answer on the basis of these data is whether the attitudes affected the test results or – just the opposite – the results shaped those attitudes.

The first stage yielded inconclusive results because its instruments did not provide a control for the variables which could affect its outcome. The crucial decision was not to use the pre-test post-test type of design. This decision was made consciously, for fear that the *practice effect* (Dörnyei 2007), resulting from taking the pre-test, could positively affect the post-test results, thus neutralizing the difference in effectiveness between the two types of instruction (corpus-based and not corpus-based). This *post-test only* design would have been successful if the two groups were indeed equal in all respects. With two groups as different as the ones involved here, the pre-test post-test research design would have prevented the problems that transpired. For each group the effect of the treatment would have been calculated on the basis of the gain between pre-test and post-test results. Even if the effect had not been visible, there would have been a point of reference for each group separately. Unfortunately, the differences became evident after the project was initiated and it was too late to introduce a pre-test.

Another problem with the design was that the questionnaire performed near the end of the experiment was anonymous, so as to allow students more freedom in expressing their opinions. Due to this, it was impossible to perform a correlational analysis that could have verified whether there was a connection between attitude and test results. If this had been the case, a co-variance analysis could have accounted for the attitude fac-

tor to some extent. This was another reason why a new attempt at the experiment had to be undertaken, where the missing elements of the design could be supplemented.

4.10. The experiment – the second stage (2010/2011)

From this point only the final stage of the study is going to be reported. The general organizing principle is the same as in the academic year 2009/2010 (see Table 15, p.242). Two major modifications were made, however. First of all, the pre-test was added to establish the point of departure for each of the groups in the study, so that the post test could measure the progress students made from that point. Second, the survey that students respond to was no longer anonymous, which makes it possible to include a correlational analysis in the set of research instruments for the study.

4.10.1. Pre-test and selection of items for the experiment

The analyses of students' examination essays and of the blog corpus material produced a body of error data, on which further stages of the study were to be based. As stated at the beginning of this chapter, the major point of the project was to test corpus-based teaching techniques in the classroom to find out how effective they are and how students perceive them. In order to achieve this, the teaching needed to address areas of language which were problematic for a significant proportion of students and which, at the same time, were relatively common in their language production in a variety of the types of discourse in which they engaged.

Another factor affecting the choice of target items was that the experimental lessons were part of the third year practical English grammar course and so they needed to be integrated into the course plan. One of the objectives of the course was to help students develop their command of grammar to the degree that would allow them to pass the final practical English examination. Therefore, the experimental lessons had to bear some relevance to what was tested in the exam. Since students needed to feel continuity and logical progression from one class to another, the experimental lessons had to be connected with what happened in the classes that preceded and followed them. Also,

students were supposed to be benefitting from them as much as they did from other classes in the course in terms of their command of English grammar. Due to all these requirements, the lexical element was not very strongly emphasized in the course and was mostly left to be dealt with in skills-oriented classes (reading and speaking).

On the other hand, students did realize that some elements of lexis are grammatically challenging and they actually welcomed a slight shift in focus from 'broad' rules of grammar to more local, lexically oriented grammar problems. Consequently, the goal here was to preserve a balance between different types of items, giving priority to those problems that constitute serious impediments to students' progress in English.

The preliminary selection of the experimental items was based on the results of the previous two stages of the study: essay analysis and then learner corpus analysis. The items which were to be chosen had to be prominent in terms of their item/error ratio, and at the same time could be combined into pairs (see below). The final decisions were based on pre-test results, which in all but one of the cases generated sufficiently high average error rates, and which at the same time met other requirements of the research design.

The pre-test, post-test and delayed test are essential elements of the design of this research project. The elicitation procedure applied in the part of the study where the effectiveness of corpus-based instruction was tested was experimental in nature. This means, according to Corder ([1976] 1981), that students were to generate particular linguistic items, which were predefined by the design of the study and the interest of the researcher. Here students' grasp of particular language items needed to be verified in experimental conditions, i.e. in a test situation. Unlike the blog corpus data, which were produced by students writing freely on very general topics, test data were produced in strictly controlled conditions and students' output was limited in terms of both content and form. Admittedly, item-based tests do not render a perfectly accurate representation of students' general language proficiency but rather scrutinize their awareness and knowledge of rules. Still, it is assumed here that the information obtained from the tests has some relevance to the changes that occur in students' command of English. Although it could not be proven unequivocally whether the items that had been taught in the experimental and control lessons became permanent elements of the students' language competence, it was possible to find out whether these lessons had an effect on the conscious choices students made when they were supposed to use the target elements of

English lexicogrammar. This was the role of the pre-test, the post-tests and the delayed tests.

There were two major aims that the pre-test was expected to achieve. The first was diagnostic: to confirm the severity of the problems occurring with the use of the items selected for the experiment. It would not have been worthwhile to examine students' progress in areas that did not constitute a challenge for them. What is more, the objectives of the study could not have been achieved if the items selected for the experiment had not actually generated high error rates. The standard pre-test/post-test design of the study required the initial result to be relatively high; otherwise no gain in accuracy could have been evidenced. This connects with the second aim of the pre-test, which was to provide a measurement tool for the progress that students were expected to make as a result of the remedial instruction provided.

For the pre-test to perform a diagnostic role in the project, its reliability, or to be more accurate, its *internal consistency*, had to be verified. The method chosen for the purpose was the commonly applied *Spearman-Brown split-half estimate* (Bachman 1990, Bachman 2004), whose advantage is that it permits a single application of a test and is relatively simple to perform. It calculates how well the halves of a test correlate with each other and then, since the primary correlation is calculated for half the length of the original test, the result is corrected for its full length. The estimate is computed according to the following formula:

$$r_{xx'} = \frac{2r_{hh'}}{1 + r_{hh'}}$$

– where $r_{hh'}$ is the correlation between the two halves of the test. The method requires two assumptions to be met: one is that the two halves should have equal means and variances, and the other that the two halves be independent of each other, i.e. that they are assessed separately and that a successful response to one item does not depend on students' performance on another (Bachman 1990). Both conditions were met in the case of the pre-test: the mean values were very close in value ($M_a = 8.69$, $SD_a = 2.32$ and $M_b = 8.21$ $SD_b = 2.24$, out of the total score of 15 for each half), which an independent samples *t*-test indeed proved not to be significantly different: $t(76) = .94$, $p = .35$ (two-tailed). The variances were $s^2_a = 5.38$ and $s^2_b = 5.01$, for which no significant difference

was evidenced in Levene's test (F = .098, p = .76). The other assumption, that the items should be independent of one another, was met as well.

In total, the split-half reliability coefficient calculated with the formula given above amounted to $r_{xx'}$ = .76, which, though not very impressive, is acceptable. Considering the unique character of the test – the fact that the elicitation technique employed in it was translation, which does sometimes generate unpredictable responses, and that the construct of the test was very narrowly defined – it has been assumed that the statistic may be underestimated to an extent. Other statistical estimates of the test prove it to be well balanced and properly structured. Skewness and kurtosis values were g_1 = .174, $se(g_1)$ = .378 and g_2 = -.359, $se(g_2)$ = .741 respectively – well within the commonly accepted ±2 norm for both (Bachman 2005), and well within the ratio obtained by dividing the kurtosis or skewness statistics by their respective standard error values, where the norm is again ±2 (also Bachman 2005). The mean and median are also very close in value (16.9 and 17 respectively) with SD = 4.1. Finally, normal distribution is evidenced by the Shapiro-Wilk statistic of W = .97 (p = .43), which indicates that there are no grounds for rejecting the null hypothesis that the distribution of pre-test scores is the same as the normal distribution. All this gives grounds to the assumption that the pre-test is a well constructed tool of measurement and therefore has been accepted as sufficient for the needs of this project.

The pre-test consisted of 31 sentences for translation from Polish into English, and contained the minimum of five items of the six language problems pre-selected for the experiment. Some sentences included two or, in one case, three different items so as to minimize the duration of the test and fatigue of the subjects, but each of the items was independent from the others, each of them was assessed separately and students were instructed to translate as much of each sentence as they could. This is important to point out, as otherwise the conditions of the Spearman-Brown split-halves estimate would not have been fulfilled. As it is stated below, some items that had originally been part of the test proved to be of no use to the project, other than serving as distractors. These were excluded from further stages of the study and, consequently, from the reliability test. The following paragraphs explain the details of how the reliability test was performed.

Thirty items of the test were grouped into six sets, each set holding five items that tested one language problem. Within these sets the items were then ranked by their

facility values (average scores which the items generated in the population sample), and then alternately assigned to one of the halves of the test (*a* or *b*), so that the split-half estimate could be calculated. This sorting principle was devised instead of random selection, which is sometimes recommended, so as to make sure that each language problem was represented by two or three items in each half. Such operations have been recommended in Bachman (2004: 161): "To minimize these potential problems, it may be preferable in some tests to use a rational, rather than a random split. That is, it might be better to decide how to split the test into halves that are equivalent in content, or what they measure". If a random selection (e.g. odd/even split) had been applied, the reliability estimate of the test would not have had much connection with what the test was supposed to measure: the degree to which the eight predefined areas of English grammar and lexis constitute a challenge to Polish advanced learners. Table 19 illustrates the process of selection described above.

Table 19. Allocation of pre-test items to parallel halves

Sentence No	Item ID	Facility value	Allocation to test halves
14	a/some 2	0.821	a
26	a/some 5	0.769	b
15	a/some 3	0.667	a
9	a/some 1	0.590	b
19	a/some 4	0.590	a
2	possibility 1	0.795	b
6	possibility 2	0.769	a
28	possibility 5	0.590	b
17	possibility 3	0.385	a
21	possibility 4	0.256	b
10	approve 2	0.590	a
24	approve 4	0.487	b
29	approve 5	0.436	a
1	approve 1	0.282	b
13	approve 3	0.282	a
22	however 4	0.872	b
12	however 1	0.641	a
16	however 2	0.462	b
18	however 3	0.462	a
25	however 5	0.462	b
23	number/amount 4	0.872	a
8	number/amount 3	0.641	b
6	number/amount 1	0.615	a
30	number/amount 6	0.590	b
7	number/amount 2	0.538	a
3	state 1	0.667	b
27	state 4	0.641	a
5	state 2	0.538	b
31	state 5	0.385	a
20	state 3	0.205	b

Translation as the elicitation technique for the pre-test (and post-tests) was chosen over others for several reasons. First of all, it allows students to consciously monitor interference errors, provided that they have had enough relevant input and translation practice in class. The technique has other strong benefits: it generates language production data from subjects of the study and at the same time, unlike open-ended writing tasks, translation allows very little room for maneuver in terms of content and form, and limits the use of avoidance strategies. Students are made to express a particular concept

or message, which reflects more closely how well they have actually mastered the use of a given target form. In this way an obligatory context for the grammatical and lexical rules investigated here could be created. Admittedly, elicitation through translation does have drawbacks: the range of possible responses is wider than in more closed-ended tasks, such as multiple choice questions or acceptability judgments, which are easier to standardize as no ambiguity in marking can arise. For the needs of this study, however, it was important that students provided samples of their own production in which they confront language problems that were of interest to the researcher. In the circumstances, translation seemed the most natural and authentic technique to choose, and the most valid in view of the research questions posed for the study.

Apart from the items included in the final version of the study, there was one which featured strongly in the essay and corpus data, but the pre-test proved it to be of little value. The problem in question was the use of a comma in front of *that* – mainly in noun clauses, but also in restrictive relative clauses and, most unexpectedly perhaps, in clauses where *that* is a demonstrative pronoun in the role of a subject. The latter is sometimes referred to as a *comma splice* – a sentence in which clauses are not properly subordinated or coordinated, but a comma is put between them instead. Such sentences are not considered acceptable in standard written English. The two target rules involved were the following:

- do not use a comma in front of *that* as a conjunction or relative pronoun unless after an embedded phrase;
- use co-ordination or subordination devices rather than a comma alone to combine clauses into sentences.

For some reason, this type of error did not feature very strongly in the pre-test. Out of 11 items that tested the above rules, the average rate of error for the whole population of students in the study was merely 5.1 percent, even though learner corpus search results indicated a much higher error rate (out of 131 occurrences of comma + *that*, 36 were incorrect). Admittedly, the statistic may not have appeared so high if overall frequency of *that* in the corpus is considered: there are 3161 tokens of *that* in the corpus, which means that only 1.1 percent of its occurrences involve the comma+*that* error. Even though not all of the occurrences of *that* appear in an obligatory context for the target rules, the scale of the problem does not appear to be substantial. One way or another, the pre-test results eliminated the problem as a potential experimental item.

The reasons the error does not appear in the translation test as frequently as it does in blog posts could be that students may be aware of the different rules in English, but were not careful enough when writing on their blogs to avoid these errors. In a translation test, where Polish-English differences are in focus, their monitor for this particular feature of English – so distinctly different from Polish – may have been higher than in the other, more relaxed form of writing.

Otherwise, the pre-test gave relatively unsurprising results, documenting a significant level of inaccuracy in the tested areas. Table 20 presents the error frequency rates for each item included in the study. The rates were calculated as follows: a percentage rate was calculated for each student by dividing the number of target errors he/she committed by five – the number of test items on a given language problem; then the mean value for all participants was calculated for each problem. The resulting rate of error is believed to represent the level of difficulty the students encounter with a given language problem most accurately.

The thus calculated error rates for individual students underwent further analysis. The mean value for all the participants of the study was used to verify the usability of particular items in the project. More importantly, after the lessons had been taught and post-tests carried out, individual students' gain values for the 5-item sets were calculated by subtracting the pre-test results from the post-test results.

Based on the considerations above, the items chosen for the second experimental attempt were as follows:

Table 20. Items selected for experimental lessons

ITEM	PROBLEM AREA	Mean rate of error (pre-test)	No of tokens/ No of errors (corpus)
however/although	syntax	42.1%	55/17
possibility/ability/opportunity	syntax (+ word choice)	44.1%	13/7
state/find/claim	word choice	51.3%	10/7
approve/approve of	word choice	58.5%	9/4
amount/number	countability	34.9%	12/5
some/a + singular countable noun	countability	31.3%	50/35

The labels in the second column indicate that the lessons are matched into three pairs. This was done to accommodate to the research plan presented earlier, which required that the lessons be combined into pairs of problems from the same or similar

areas of language and of similar difficulty, so that some comparisons between these lessons could be made.

It was also important that the students had similar learning experiences to relate to at the end of the experiment, because they were expected to express their opinions on the corpus-based lessons in a follow-up survey. For these opinions to be comparable between groups, the corpus based lessons had to be similar in their content and level of difficulty.

4.10.2. Experimental lessons

The most important element of the study was teaching the selected language problems in experimental lessons with the use of various corpus-based techniques and materials, such as analysis and sorting of simple concordances, completion of gapped concordances, translation or matching based on parallel corpus search results, concordance-based partial translation, error correction based on learner corpus concordances, and activities involving word or phrase frequency lists. All the corpus materials were used in paper format and students did not perform corpus searches in the classroom, so as to provide students with exactly the same input and minimize the number of variables involved in the experiment. Although the original idea of the DDL was to provide learners with corpus tools and allow them to formulate their own searches, printed materials were chosen as more practical in terms of the design of the experiment and more feasible in the context of Polish educational facilities. For the sake of the experiment, it was important to minimize the variety of input among the participants. Otherwise analysis of the effectiveness of the techniques would have been challenged by an uncontrolled variable – the quantity and quality of data a given student actually managed to process during the experimental lessons. For these reasons, only paper-based concordances were used in the current study. Apart from these lessons, however, the students were offered instruction in using corpora for language reference, and some of them later reported to be using them regularly outside the classroom.

As far as adapting to the educational facilities is concerned, one of the study's aims was to prove that corpus tools are accessible and usable in an average English lesson in a Polish school, so assuming that every student could use an online computer in

class would be hardly realistic. Even though there are computer labs in most Polish schools, these are not normally available for lessons other than information technology, and even if they are, this requires special arrangements and planning well ahead of the lesson. The activities proposed and analyzed here were not meant to be special occasions, but regular elements of English lessons in a standard environment. Also, such an arrangement prevented any confusion that might result from operating concordancers, which requires a degree of technical skill (and therefore training) and which might not be very appealing to some students. For these reasons, all the materials used in the project were printed and brought to the classroom for students to analyze.

For the sake of control in the experiment, parallel classes were taught with conventional forms of foreign language instruction. These involved Polish and English sentence analysis, sentence building, sentence combining and transformation, sentence recall, translation (L1 > L2 and/or L2 > L1), story writing, error correction, and word maps. Care was taken to make both sets of lessons attractive and rich in content, so that no bias in any direction would occur. Extensive lesson reports which summarize the development of each experimental and control lesson are placed in Appendix D. Below is a list of the lesson topics:

(1) *however* – used as a subordinating conjunction of concession, which it is not;
(2) *possibility* – confused with *opportunity*, *ability*, *power*, etc.; complementation problems;
(3) *state* – confused with *find*, *claim*, *remark*, *say*, etc.
(4) *approve* + noun – confused with *approve of* + *noun*;
(5) *a* / *some* – *some* overused in front of singular countable nouns;
(6) *amount of* / *number of* – *amount of* misused with plural nouns

4.10.3. Pre-test / post-test gains analyses

Every lesson that was part of the experiment was followed by a post-test a week later. The test included those sentences from the pre-test that contained items presented in the experimental or control lesson the week before. The sentences were slightly modified so as to limit the practice effect in the students (Dörnyei 2007). Below is an analysis and discussion of the results for lessons on the same lexicogrammatical problem, paired for

the sake of comparison between the DDL techniques and the generally accepted, conventional ones. For each lesson a gain in individual students' results was calculated by subtracting the number of errors in the post-test from those in the pre-test. The gain value represents the degree to which the number of errors was reduced from an expectedly higher value before the treatment to a lower one afterwards. The higher the gain value then, the better the effectiveness of the lesson for a particular student (fewer errors committed in the post-test in comparison with the pre-test). As can be seen from the data, the outcome was not always like this, hence instances of negative values. Sets of thus calculated gains were compared for pairs of corpus-based and conventional lessons. In each case a different number of students were involved, because only the results of students attending the experimental and control lessons were included in the study. After that, an overall analysis is presented for all the experimental and control lessons. All the results of the pre-test, post-tests and delayed tests are placed in Appendix D.

4.10.3.1. HOWEVER

The lesson focused on using the linking adverb *however* and distinguishing it from the subordinating conjunction *although*. A common error of Polish students is using *however* in place of *although* and *though*, which can be attributed to the dual function of *jednak*, *jednakże*, *aczkolwiek*, etc. in Polish grammar: each of these words can be used as either a conjunction or a linking adverb. More discussion of this error can be found in section 4.6.3.3. p. 217, and section 4.8.3. p. 239. For lesson reports and class materials see pp. 396-400.

Table 21. General statistics for lessons on *HOWEVER*

	TECHNIQUE		Statistic	Std. Error
GAIN	CORPUS-BASED (gr 1+2)	N	19	
		Group	1+2	
		Mean	0.32	0.37
		Median	0.00	
		Variance	2.56	
		Std. Deviation	1.60	
		Minimum	-3.00	
		Maximum	3.00	
		Skewness	-0.40	0.52
		Kurtosis	-0.39	1.01
	CONVENTIONAL (gr 3)	N	16	
		Group	3	
		Mean	0.88	0.52
		Median	1.00	
		Variance	4.25	
		Std. Deviation	2.06	
		Minimum	-4.00	
		Maximum	4.00	
		Skewness	-0.75	0.56
		Kurtosis	0.74	1.09

Data obtained for the lessons given in Table 21 show that the mean as well as the median of the gain for the conventional lesson are higher than for the corpus-based lesson. The range of results is broader for that lesson as well, both upwards and downwards (-4:4 vs. -3:3). Judging from the kurtosis and skewness values and the relative proximity of the mean and median values for each group, both sets of results seem to have relatively normal distributions, which has been confirmed in the Shapiro-Wilk test of normality: $W_{corp} = .95$ ($p = .42$) and $W_{conv} = .947$ ($p = .45$). Thanks to this, conditions are met for an independent samples test for the equality of means to be applied in order to establish whether the differences in gains that occurred in the two groups have a statistical significance, i.e. are likely to result from the difference in the treatment rather than from chance. Levene's test gives no grounds for rejecting the assumption of equality of variance, and so a t-test can be applied. Its results are as follows: $t(33) = -.90$ ($p = .37$). Since the significance value for the *gain* is well above 0.05, it must be concluded that the difference between the two means is not statistically significant, which means it can be an effect of chance (the more conservative 2-tailed significance value has been considered).

Figure 18 below combines a lot of the information presented in Table 21 above: the medians (0 and 1), the fairly normal distribution, the slightly better results for the

conventional lesson, and the fact that the difference between the two groups is not statistically significant. There is also an outlier (the "-4" value in the 'CONVENTIONAL' group). It is marked as '3b', a symbol representing a student with an uncharacteristically low, negative mark (incidentally, the student took part in the first two lessons only, and later abandoned his studies).

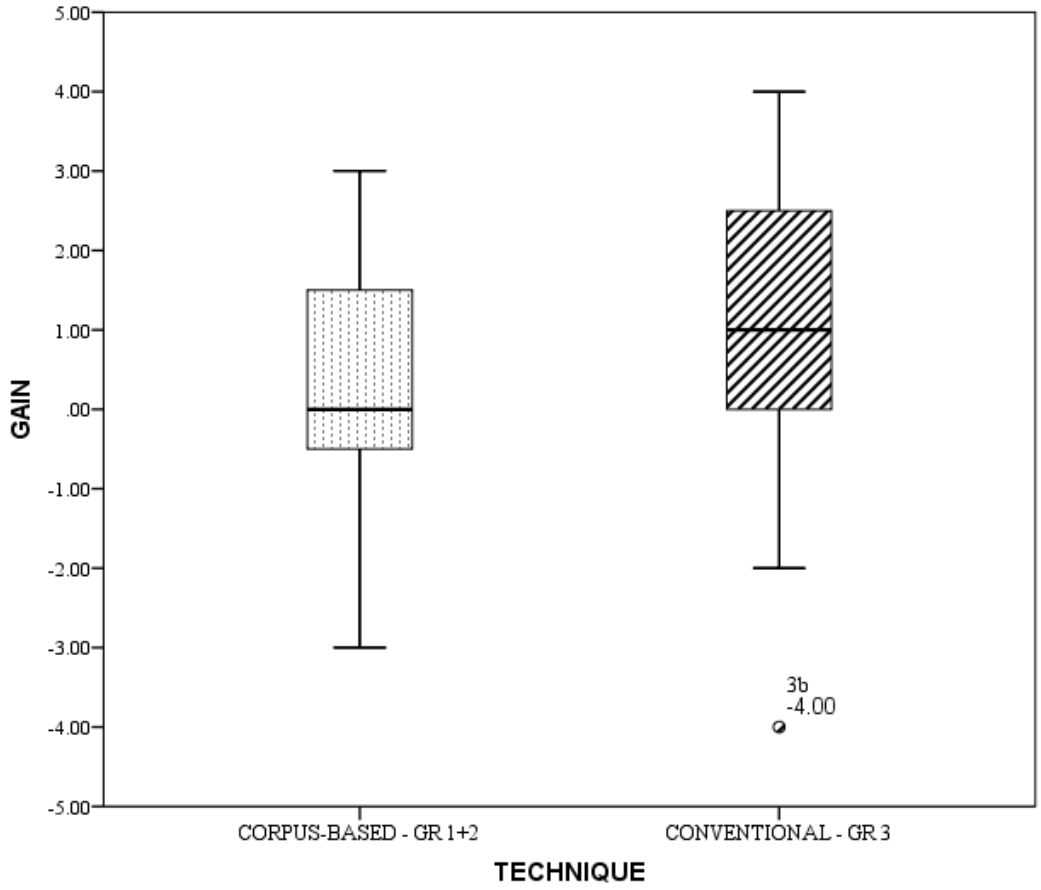


Figure 18. Boxplots for lessons on *HOWEVER* – pre-test/post-test gains

What is most disappointing perhaps is how low the values of gains in both groups are. Even for the conventional lesson, which seems to have been slightly more successful, the mean is not very impressive (M_{conv} = .88). The results could mean that students did not really learn much from the lessons and made little progress. The two groups received short feedback 'sessions' on the post-test and, a month later, another test was applied, which will be referred to as a *delayed test* here. Its purpose was to check whether there was any change in the students' grasp of the problem after some time, and if so, how the two groups compared in this respect. The gain results of this

test calculated against the pre-test again are presented in Figure 19 – together with those for the post-test. It must be noted that the N values for the delayed test are lower than for the post-test (17 and 13), as not all the students who attended the experimental/control lesson were available on the day of the delayed test. The means and medians obtained for gains in the delayed test are higher: $M_{corp} = 1.18$, $SD_{corp} = 1.77$ ($m_{corp} = 1.00$) for the corpus-based class and $M_{conv} = 1.92$, $SD_{conv} = 1.44$ ($m_{conv} = 2.0$) for the conventional class. Normally students would be expected to perform slightly worse after a passage of time, but considering that the initial results were rather poor and that the first test provided another opportunity to revise the problem, the outcome is not very surprising.

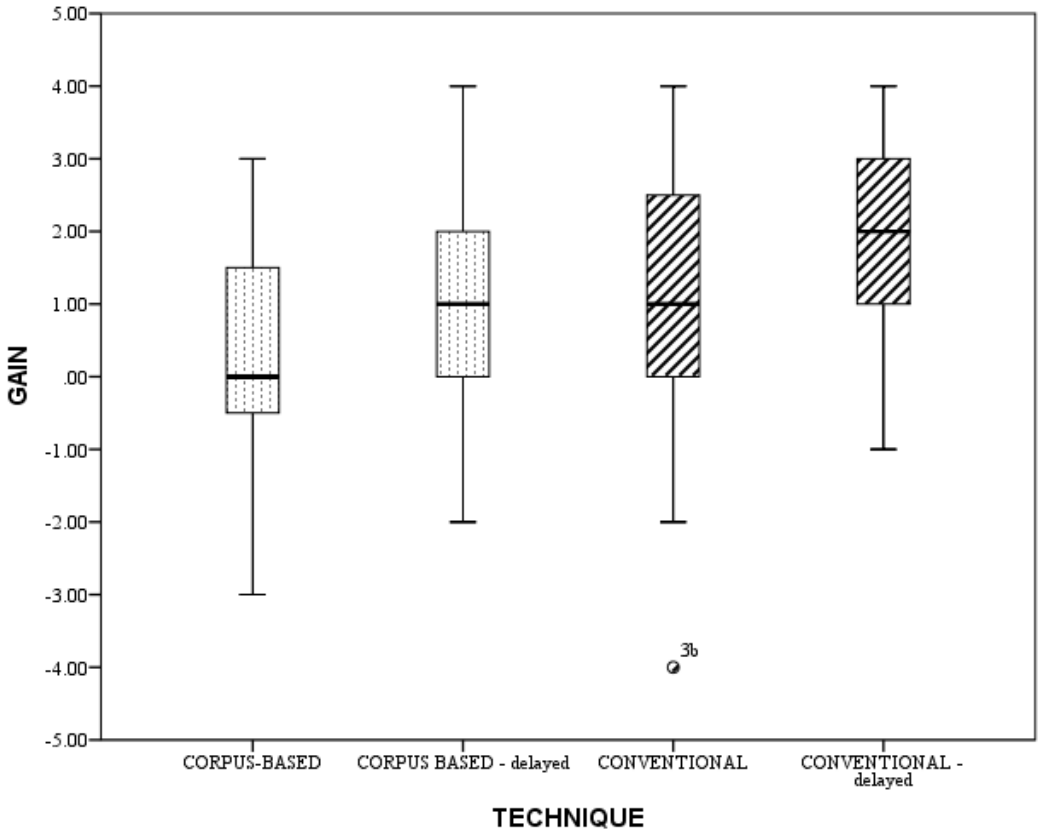


Figure 19. Boxplots for lessons on *HOWEVER* –
pre-test/post-test gains and pre-test/delayed test gains

As the graph demonstrates, each of the groups has made a similar improvement over time. Still, the corpus group's gain value remains lower than the other one's. Re-

sults of statistical analyses for the delayed test were very similar to those for the post-test: normal distribution, similarity of variance, and an independent t-test result indicating no significant difference between the means ($t(28) = -1.24$ and $p = .23$).

Another statistical test was performed to examine the significance of the change in gain values resulting from the time factor combined with the important fact of students having received feedback on the post-test. It transpired that this factor was significant for the corpus-based group: a paired-samples t-test was performed for the post-test and delayed test gain values: $M_{post} = 0.24$, $SD_{post} = 1.64$ and $M_{del} = 1.18$, $SD_{del} = 1.77$; $t(16) = -2.32$, $p = .034$, (two-tailed), $d = .56$. According to Cohen (1988) this effect size (d) may be judged as medium. (The statistics for the post-test are slightly different from those in Table 21 because for a few participants there were no data available for the delayed test.) As a result, the null hypothesis of equality of means for the two tests could be rejected, and the alternative hypothesis was accepted, according to which test results for the delayed test were significantly higher than for the post-test. As Cohen's d-value indicates, the difference is not large, but notable. A possible explanation is that the test itself and/or the teacher's feedback had a significant effect on the final outcome of the experiment. With the other group the difference of means was a little smaller ($M_{pre} = 1.15$, $SD_{pre} = 1.68$ and $M_{del} = 1.92$, $SD_{del} = 1.44$), and so the t-test did not show it to be significant: $t(12) = -1.64$, $p = .13$ (two-tailed). In order to avoid such an unplanned effect in the lessons that followed, the instructor decided not to offer students feedback on the post-tests that were to be administered after the subsequent lessons, until the delayed tests had been performed.

Analysis of data for the first lesson was not very encouraging: the corpus lesson was found to be slightly less effective, although the difference did not prove to be of statistical significance. This outcome may result from one characteristic feature of the lesson – the fact that its subject matter involved a syntactic problem, the processing of which may require a broader context than a concordance can provide. Some of the examples used in the experimental lesson were fragmentary and/or difficult to understand without a broader context. Inauthentic examples in the control lesson, though artificial, may have been easier for students to process and to build upon. Judging from the first lesson of the project then, corpus-based teaching may not prove such a successful technique as had been previously expected. A possible explanation of the outcome might be that the corpus-based lesson was too much of a novelty for students, and they may have

needed more time and experience for this type of lesson to be effective. Analyses of the lessons that follow should throw more light onto these questions.

4.10.3.2. POSSIBILITY

The lessons that were paired with the ones on *however* were devoted to another common Polish-English difficulty: the use of the word *possibility*. The problem is quite complex, because there are actually two choices involved: lexical and syntactic. In both cases Polish seems to interfere: the Polish equivalent of *possibility* is *możliwość*, but the Polish word has a much broader range of meanings, which overlaps with *possibility* only partially. A more detailed discussion of the error can be found on page 233 and in Appendix D, p. 410, together with all class materials. The same error was quoted in Kaszubski (1999) as a typical feature of Polish learners' English.

The control lesson on *possibility* was designed in such a way as to eliminate the element of novelty in the comparison with the corpus-based technique: while the other group had a corpus based lesson on *possibility*, the control group was given a lesson that employed word maps, also a new teaching instrument, and one that students are believed to find attractive and interesting. The technique must have been familiar from other courses, but it had not been used in the third year grammar course before, and so might heve been perceived as a novelty.

From the pre-test/post-test gains presented in Table 22 and Figure 20 below, it can be concluded that indeed the word maps proved a much more effective instrument, at least in this particular case.

Table 22. General statistics for lessons on *POSSIBILITY*

	TECHNIQUE		Statistic	Std. Error
GAIN	CORPUS-BASED (gr 3)	N	16	
		Group	3	
		Mean	-0.06	0.28
		Median	0.00	
		Variance	1.26	
		Std. Deviation	1.12	
		Minimum	-3.00	
		Maximum	2.00	
		Skewness	-0.83	0.56
		Kurtosis	2.46	1.09
	CONVENTIONAL (gr 1+2)	N	19	
		Group	1+2	
		Mean	1.42	0.37
		Median	2.00	
		Variance	2.59	
		Std. Deviation	1.61	
		Minimum	-1.00	
		Maximum	4.00	
		Skewness	-0.06	0.52
		Kurtosis	-0.99	1.01

The most striking difference between the two sets of data is the most important one: the difference in means and medians: $M_{corp} = -0.06$, $SD_{corp} = 1.1$ ($m_{corp} = 0$) and $M_{conv} = 1.42$, $SD_{conv} = 1.6$, ($m_{conv} = 2$). The gains for the corpus-based lesson have a distribution that is significantly different from the normal one. The negative value of skewness is not problematic, considering its ratio to standard error, but the kurtosis is very high, which means that the distribution has a leptokurtic shape, with most results highly concentrated around the mean. When considered in combination, the values for the kurtosis and mean indicate here that a large majority of students did not benefit from the corpus-based lesson at all, and quite a few were actually confused (hence the negative gain values).

The lesson with word maps, on the other hand, proved to be quite successful, especially when compared with the other lesson. As the Shapiro-Wilk test of normality confirmed that the gain score distribution for the corpus-based class was not normal (W = .878, p = .036), the difference between the two groups was tested in a nonparametric equivalent of an independent samples *t*-test, i.e. the Mann-Whitney U test, where $U(35) = 74.5$, p = .009 (two-tailed), r = .33. The test evidenced a statistically significant difference in the gain scores for the corpus based and word-map based lessons on *possibility*, indicating that the former was in fact significantly less effective. The last statistic (the strength of association) informs about the effect size, and in this case could be interpreted as having a *medium* effect (Cohen 1988).

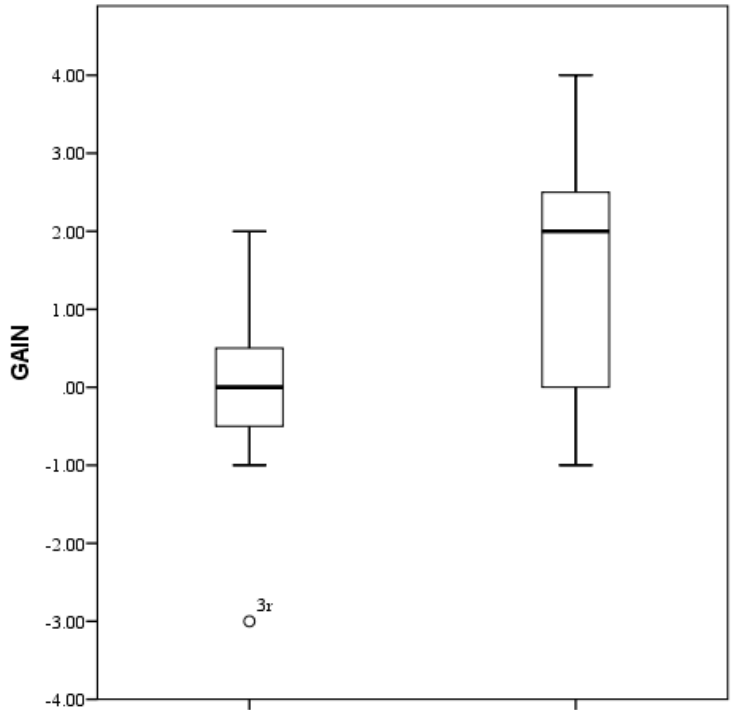


Figure 20. Boxplots for lessons on *POSSIBILITY*

The results of the delayed test for *possibility* (presented in Figure 21) are quite different from those on *however*. There was a change over time in both groups, but it was different for each of them. The distribution of results for the corpus group was no longer leptokurtic but changed to normal; the mean and median remained the same (M_{corp} = 0, SD_{corp} = 1.1, m_{corp} = 0), so there was no delayed increase in gain, such as had occurred for *however*. It seems that in the case of *possibility* neither the post-test nor the feedback that students received afterwards had really any favorable effect on the results. As for the word-map lesson, its initial success appears to have worn off slightly after a month and what looked like a really good result became less impressive, though it was still better than that for the corpus group (M_{conv} = 0.82, SD_{conv} = 1.4, m_{conv} = 1).

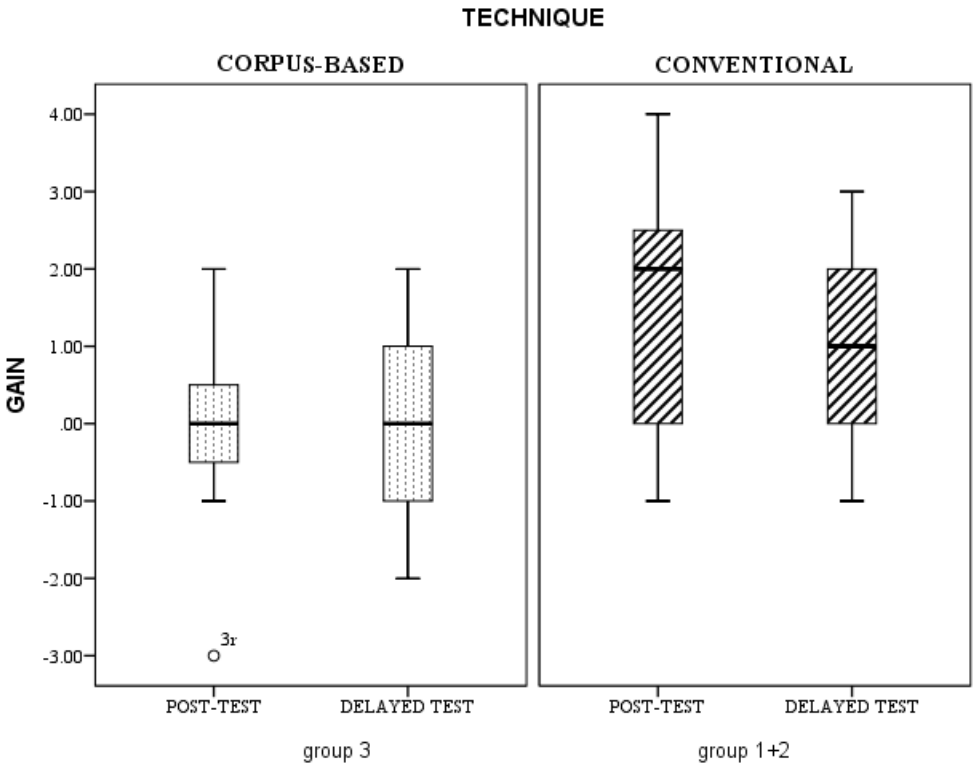


Figure 21. Boxplots for lessons on *POSSIBILITY* –
pre-test/post-test gains and pre-test/delayed test gains

After the normality of distribution and equality of variance were confirmed, a paired-samples t-test was conducted to compare the post-test and delayed test gain values for the conventional lesson group, so as to establish whether the time factor had a statistically significant effect on how much students benefitted from the word-map les-

son on *possibility*. There was a significant difference in the scores: $t(16) = 2.6$, $p = .018$, $d = .64$. This means that the positive effect of the conventional lesson was short-lived, and was reduced after one month, with a medium effect size. The corpus-group's results were not tested as the means and medians did not show any change over time.

In reaction to the above conclusion, another question arose: Is the difference between the two groups still significant after a month's time? A null hypothesis was formulated that the delayed test gain results would not differ significantly for the corpus-based and conventional lessons. The statistics were as follows: $M_{corp} = 0$, $SD_{corp} = 1.08$ and $M_{conv} = .82$, $SD_{conv} = 1.38$. As it can be concluded from the boxplot above, the distribution of both groups' results is normal; also, Levene's test for equality of variance gave positive results ($F = 2.07$, $p = .16$). Therefore, an independent samples t-test was chosen to verify the hypothesis and its results failed to find sufficient evidence for the null hypothesis to be rejected: $t(28) = -1.77$, $p = .087$ (two-tailed). The result does not fit within the confidence interval assumed for the study (.95), although it must be said that it would be acceptable if slightly less conservative criteria had been assumed (.90). As it is, the t-test did not prove with satisfactory confidence that the two groups' gain values in the delayed test were different. After all, the two classroom techniques did not prove to be very different as far as their long-term effectiveness is concerned.

4.10.3.3. STATE

The lessons on using the verb *to state* were expected to pose some difficulty for students as the pre-test error rate for the relevant items was rather high (51.3 %). The problem students have with the word is that it is very often perceived to be an equivalent of the Polish *stwierdzić*, which, like the previously discussed *możliwość*, has a wider range of meanings and is stylistically more neutral than the very formal *to state* in English. A more extensive discussion of the problem can be found in section 4.8.2. , p. 238. The error seems to occur most often in translation, but the author of this study has observed its high incidence in a gap-filling task as well. The lexical character of the problem, however, indicated that corpus materials should be particularly useful in dealing with the problem, as it is usually vocabulary learning that is associated with the use of con-

cordances. All teaching materials for the two lessons can be found in Appendix D, pp. 413-415.

Table 23. General statistics for lessons on *STATE*

	TECHNIQUE		Statistic	Std. Error
GAIN	CORPUS-BASED (gr 1+2)	N	16	
		Mean	2.00	0.30
		Median	2.00	
		Variance	1.47	
		Std. Deviation	1.21	
		Minimum	0.00	
		Maximum	4.00	
		Skewness	0.00	0.56
		Kurtosis	-0.47	1.09
	CONVENTIONAL (gr 3)	N	14	
		Mean	1.36	0.44
		Median	1.00	
		Variance	2.71	
		Std. Deviation	1.65	
		Minimum	-1.00	
		Maximum	4.00	
		Skewness	0.05	0.60
		Kurtosis	-1.34	1.15

The expectations concerning the outcome of the corpus-based lesson proved to be at least partly correct. Students' test results showed much improvement compared with the pre-test (see Table 23 above for details). The mean gain value was the highest so far ($M_{corp} = 2$, $SD_{corp} = 1.21$, $m_{corp} = 2$) and higher than in the control group ($M_{conv} = 1.36$, $SD_{conv} = 1.65$, $m_{conv} = 1$). The results were tested for their normality of distribution with Shapiro-Wilk tests ($W_{corp} = .929$, $p_{corp} = .234$ and $W_{conv} = .915$, $p_{conv} = .186$), which indicated that the distributions were sufficiently close to normal. Levene's test for the homogeneity of variances also gave positive results, though the significance value here was very close to the accepted threshold of .05 ($p = .074$), below which the null hypothesis of equality of variances would have had to be rejected. With the results generally confirming the parametric quality of the data, an independent samples t-test was performed, according to which the difference between the means for the two groups was not significant ($t(28) = 1.229$, $p = .229$). The result corresponds to what can be seen in Figure 22 below: although there is a visible difference in the median, the upper bounds of the distribution for both sets of results reach the same values, and the interquartile ranges overlap a great deal. To summarize, both techniques were success-

ful, and although the corpus based lesson gave slightly higher results, the difference may have resulted from chance rather than from features of the treatment itself.

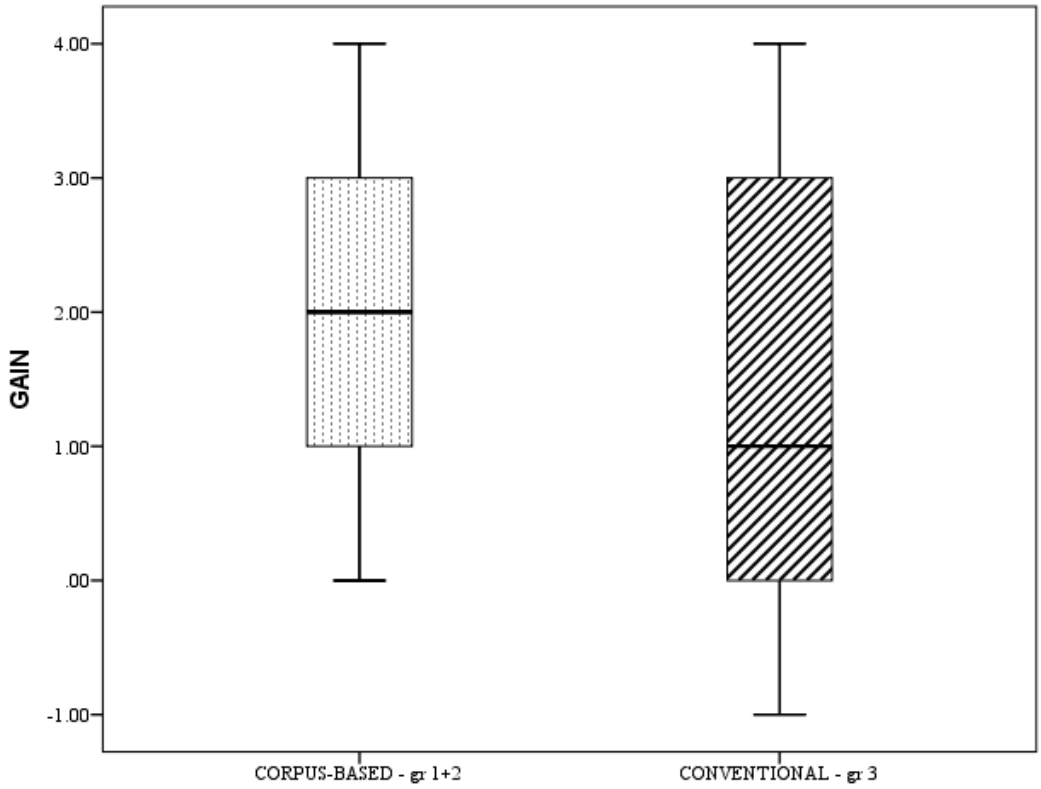


Figure 22. Boxplots for lessons on *STATE*

As previously, the effect was measured again in a delayed test a month later. This time students received no feedback on the post-test, so that the effect of the treatment itself could be assessed, without other factors obscuring the overall picture. The outcome was quite unexpected: while the median remained unchanged for the corpus-based group, its mean and distribution took a slight downward shift (see Figure 23): $M_{corp} = 1.47$, $SD_{corp} = 1.06$; $m_{corp} = 2$. The conventional group's results, on the other hand, improved: $M_{conv} = 1.69$, $SD_{conv} = 1.49$, $m_{conv} = 2$. Eventually, the two treatments produced similar means, though with some differences in distribution, especially below the first quartile (the lowest 25% of scores), which reached much lower values in the control group. As the Shapiro-Wilk test showed, the distribution of delayed test results in both groups was significantly different from normal: $W_{corp} = 0.839$, $p_{corp} = 0.013$ and $W_{conv} = 0.832$, $p_{conv} = 0.017$. Therefore, an independent samples Mann-Whitney U test was performed to compare the two groups' delayed test results: $U(28) = 79.5$, $p = .39$.

Predictably, the null hypothesis of there being a significant difference in distribution of gain values across categories of technique could not be rejected. This can also be seen in the graph below:

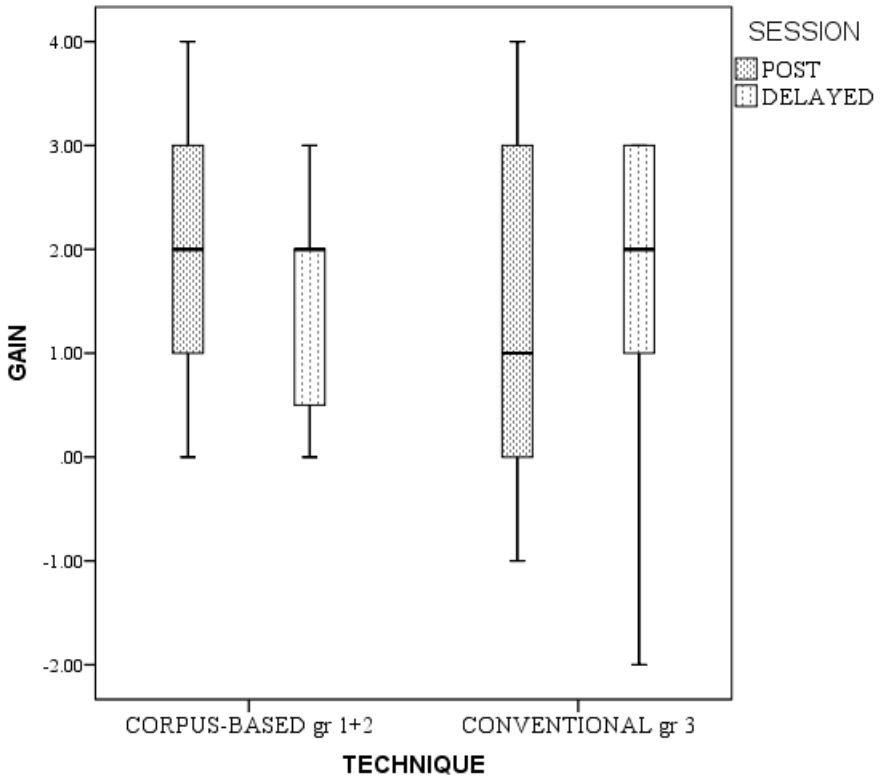


Figure 23. Boxplots for lessons on *STATE* –
pre-test/post-test gains and pre-test/delayed test gains

What is more, when the time factor was tested in a Wilcoxon's signed ranks test (i.e. a paired samples test was used to compare the post-test and delayed test gains within each group), these score adjustments did not prove to be of significant statistical value as well. All this means that in the case of these two lessons, neither the choice of technique of instruction nor delay in time had an effect on the outcome. Both lessons in which the use of the verb *to state* was presented and practiced proved relatively successful. Some minor differences between the groups were observed soon after instruction, but these were leveled after a period of one month. Students seemed to have been unaware of the error and, judging from the pre-test, many of them had committed it before the problem was raised in class. Perhaps this is why both groups showed considerable interest in the topic.

4.10.3.4. APPROVE

The lessons on the verb *to approve* were very similar both in form and in subject matter to those on *to state*. The error that was the target of the next pair of lessons was the use of the verb *to approve* with and without the preposition *of*, each of which is different in meaning and has a different set of Polish equivalents (*approve* + N – *zatwierdzić coś, wyrazić zgodę na coś* vs. *approve + of* + N – *pochwalać coś, sprzyjać komuś*). The assumed L1 influence here was the Polish learners' use of the verb without the preposition, as both its equivalents are used in Polish, for the exact meaning that would require its use (underdifferentiation).

Table 24. General statistics for lessons on *APPROVE*

	TECHNIQUE		Statistic	Std. Error
GAIN	CORPUS-BASED (gr 3)	N	14	
		Mean	2.64	0.37
		Median	3.00	
		Variance	1.94	
		Std. Deviation	1.39	
		Minimum	0	
		Maximum	4.00	
		Skewness	-0.64	0.60
		Kurtosis	-0.98	1.15
	CONVENTIONAL (gr 1+2)	N	16	
		Mean	2.63	0.42
		Median	2.50	
		Variance	2.78	
		Std. Deviation	1.67	
		Minimum	0	
		Maximum	5.00	
		Skewness	-0.003	0.56
		Kurtosis	-0.98	1.09

As Table 24 demonstrates, the gain results for the two lessons are very similar. The means are almost identical ($M_{corp} = 2.64$, $SD_{corp} = 1.39$ and $M_{conv} = 2.63$, $SD_{conv} = 1.67$), and the medians are also not far apart ($m_{corp} = 3$ and $m_{conv} = 2.5$). Since there is a difference in the medians, the two sets of results were compared to find if the difference is statistically significant. First, the basic criteria for a t-test were checked: despite the kurtosis and skewness values being close to normal, the difference between the mean and median in the corpus-based group did not indicate a normal distribution. The Shapiro-Wilk test of normality confirmed that ($W_{corp} = .85$, $p_{corp} = .02$). The histo-

gram in Figure 24 indeed shows the distribution is skewed to the left and generally does not form a clear pattern (the curve shows what shape normal distribution would be like in the dataset).

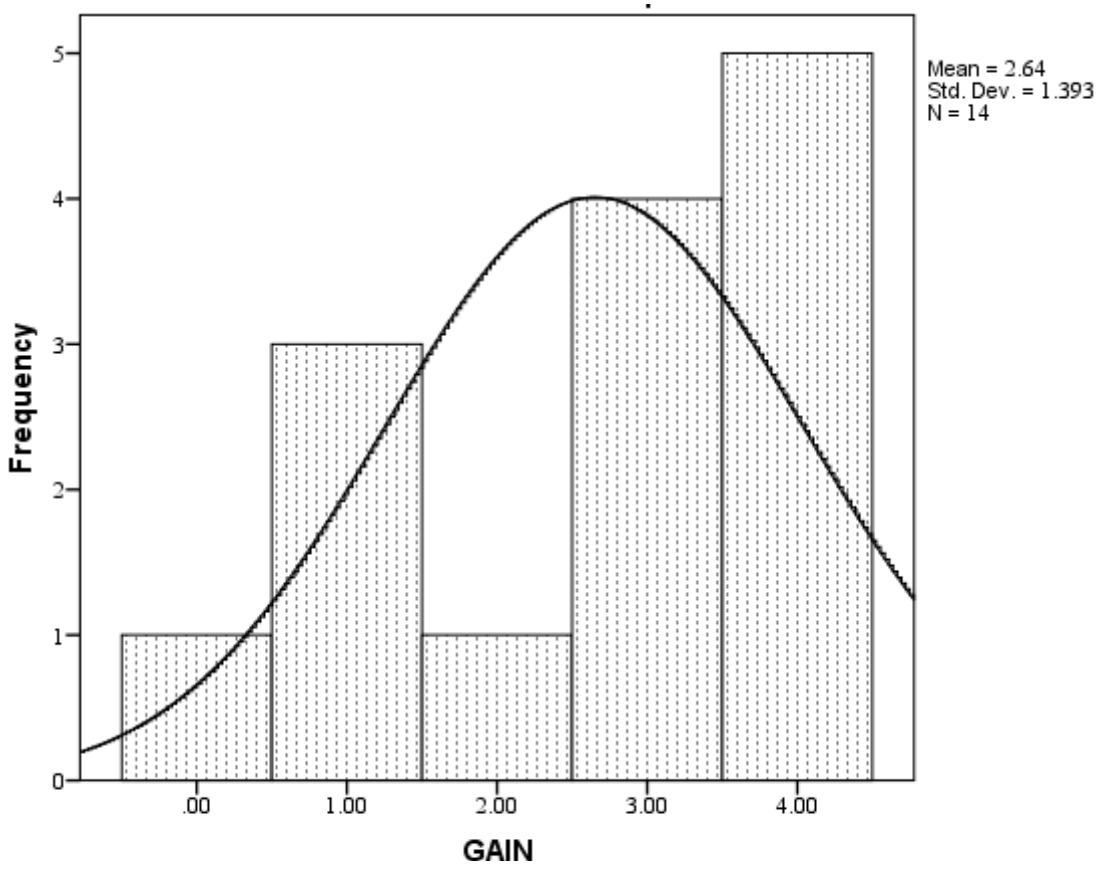


Figure 24. Post-test gains histogram for *APPROVE* (corpus-based group)

Consequently, the non-parametric Mann-Whitney U test was applied to verify the null hypothesis that the corpus-based group and the control group did not differ in their results of the post-test. The null hypothesis could not be rejected as the test statistics were very high: $U(30) = 110$, $p = .93$. This means both groups made similar progress after the lessons and the median differences between them are not statistically significant. The outcome of the test can also be seen on the boxplot in Figure 25, where the two groups' ranges of gain values overlap almost perfectly, the difference being only in the conventionally taught group's top scores reaching 5, while the corpus group's scores reached the value of 4. The small difference between the median and the third (and fourth) quartile in the corpus-based group indicates a high concentration of scores within that range (3 and 4). The same shows on the histogram above. The control

group's gain values that exceed the median are dispersed more evenly, with a wider range of values (between 3 and 5). One must bear in mind that the median value of 2.5 is a result of the even number of middle scores (2s and 3s here), in which case a mid-value is assumed; no half scores were granted in the test.

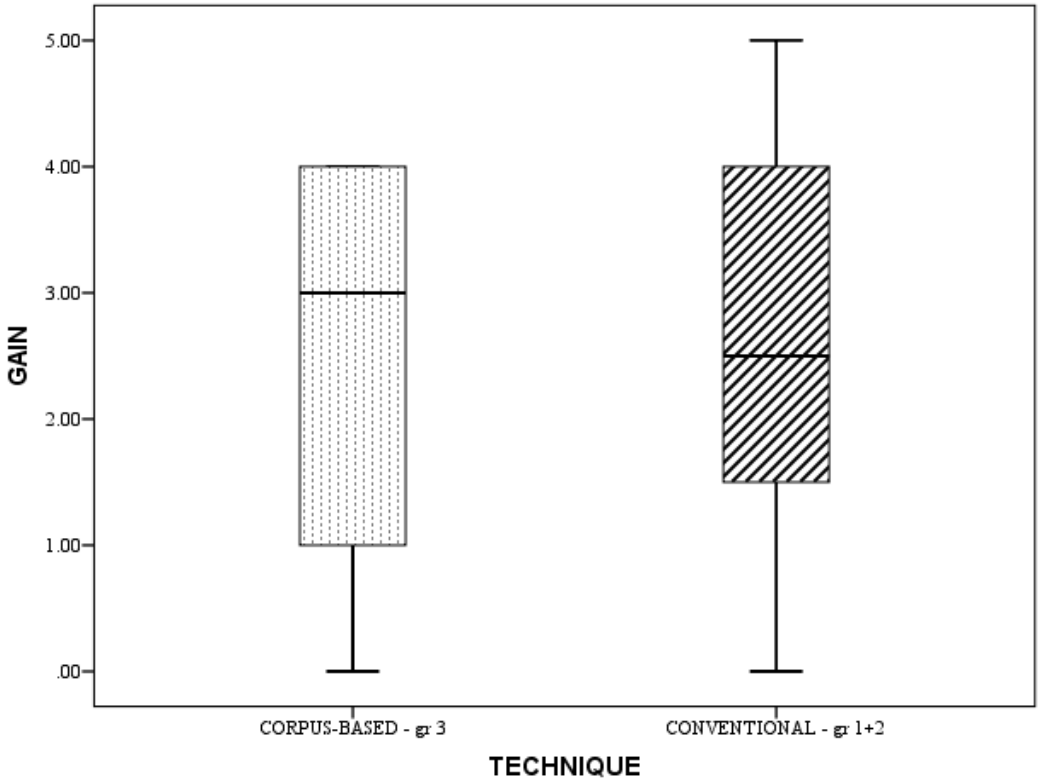


Figure 25. Boxplots for lessons on *APPROVE*

The time factor had a slightly different effect in this case than it did with the lessons on *state*. When a delayed test was applied, the improvement in scores that was observed for the conventional group in the case of *state* did not occur for *approve*. Here both groups had a slight downward shift in gain values, reaching $M_{corp} = 1.77$, $SD_{corp} = 2.05$, and $M_{conv} = 2.6$, $SD_{conv} = 1.72$. The medians for both groups also went down to 2.

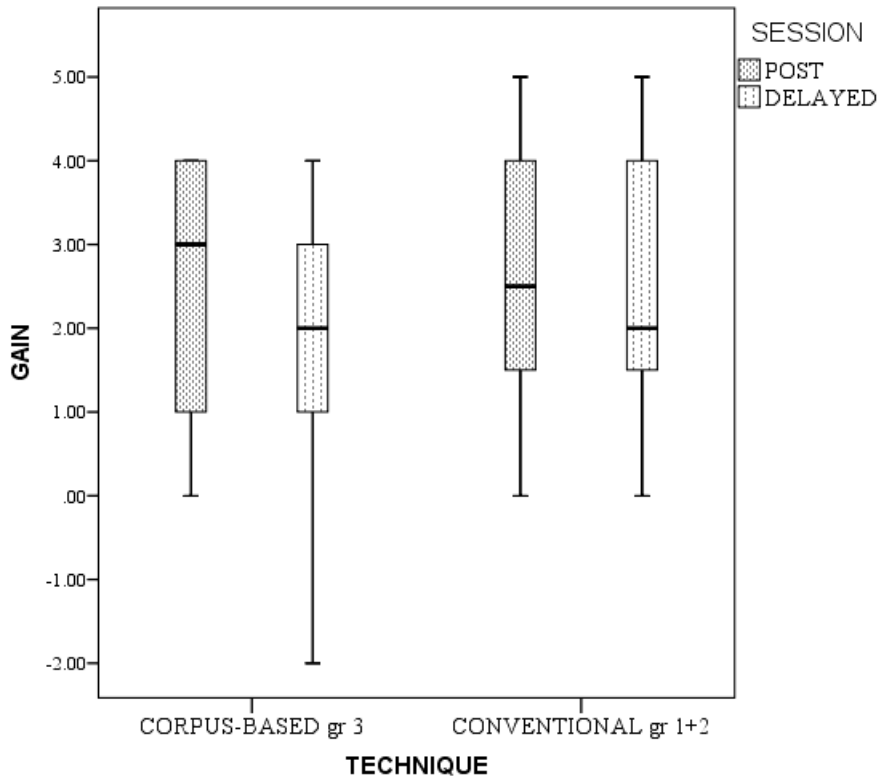


Figure 26. Boxplots for lessons on *APPROVE* – pre-test/post-test and pre-test/delayed test gains

Generally, however, the results for the conventionally instructed group did not change dramatically, except that the median shifted downwards by half a point. This is not the case in the corpus-based group, where the reduction in gain values is more visible: a couple of students' gains went as low as -2, which means they obtained scores 2 points lower than in the pre-test. Since the distribution of the corpus-based group's post- and delayed test gain values proved non-normal, the change was tested in a non-parametric paired samples test – using the Wilcoxon signed ranks test. It turned out that the difference was indeed big enough to have a statistical significance: $Z(13) = -2.04$, $p = .041$ (two-tailed), $r = .56$. The last statistic is the strength of association indicating an effect size, calculated by dividing the *z*-value by the square root of N. In this case it can be considered large (Cohen 1988). Nothing like this took place in the control group.

It can be concluded from the analysis above that the initially positive effect of the corpus-based lesson was diminished after a month's time. The effect did not occur with the conventional lesson, whose results were not significantly better than the other one's, but remained more stable over time. Since no feedback on the post-test was of-

fered to the students in the second round of the experiment, the changes between post-test and delayed test can be attributed to the passage of time, or to the very experience of writing the post-test, which can make learners verify their command of the forms tested and affect the results obtained afterwards.

4.10.3.5. SOME

The next two lesson topics are both related to an aspect of English grammar that is probably the most difficult for Polish learners: the grammatical category of countability and the choice of determiners and quantifiers that depend on it. The problem for learners is that Polish does not have an unambiguous structural or functional counterpart for this category. There are no function words or strict rules that depend on the distinction.

The first of the two lessons concerned the way of referring to an indefinite noun. In Polish indefiniteness is marked by the use of the determiner *jakiś*, both for countable and uncountable nouns. As a result, Polish learners overuse *some* in front of countable nouns, where it is normally used for emphasis and in lexical phrases like *some day*, *some kind of*, or *to some extent*. The classification of the error may be debatable, but again the broad category of lexicogrammar is the most suitable, with its sub-set of *article* as the most adequate here. In many contexts the use of *some* is not natural or expresses a meaning that is not intended by the speaker, and a neutral indefinite article *a* should be used instead, or *one* – slightly more emphatic, but suitable for singular countable nouns in some contexts. Students had had some awareness of the problem before the lesson, but since it is, admittedly, a matter of finer distinctions and the error is not grave, they did not seem to be strongly motivated to deal with it. Lesson reports and class materials can be found in Appendix D, p. 430.

From the data in Table 25 it can be expected that the gain values obtained, especially in the corpus group, would not have a normal distribution, as the means and medians are quite different in value. A Shapiro-Wilk test confirmed that observation for the corpus-based group, and hence non-parametric tests, needed to be applied for further analysis.

Table 25. General statistics for lessons on *SOME*

	TECHNIQUE		Statistic	Std. Error
some / a GAIN	CORPUS-BASED (gr 1+2)	N	17	
		Mean	0.94	0.37
		Median	0.00	
		Variance	2.31	
		Std. Deviation	1.52	
		Minimum	-1.00	
		Maximum	4.00	
		Skewness	0.84	0.55
		Kurtosis	-0.001	1.06
	CONVENTIONAL (gr 3)	N	17	
		Mean	0.82	0.39
		Median	0.00	
		Variance	2.53	
		Std. Deviation	1.59	
		Minimum	-2.00	
		Maximum	4.00	
		Skewness	0.64	0.55
		Kurtosis	0.43	1.06

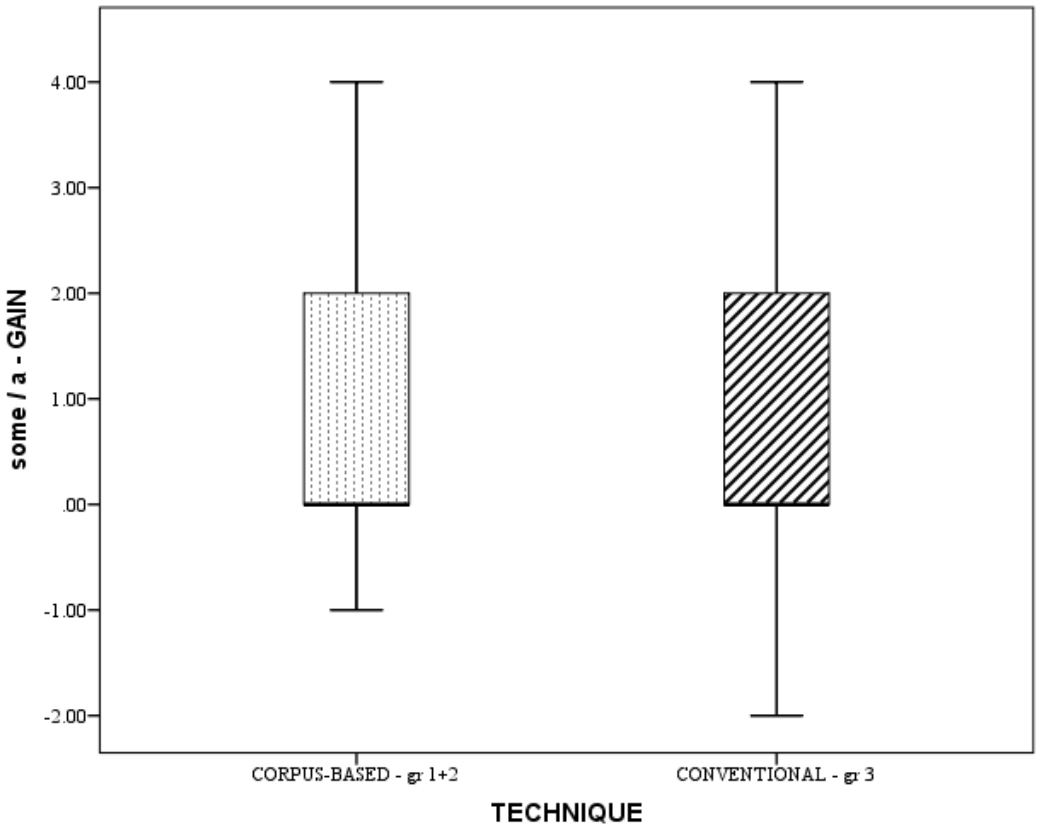


Figure 27. Boxplots for lessons on *A* / *SOME* (gain values)

It can be seen from Figure 27 that the data for the two groups are almost identical, so the statistical test for a difference of means (or medians) is not needed – there is no doubt that the distributions are very similar. However, this similarity, as well as the zero-values of the medians for gains, caused some interest in whether pre-test and post-test results were also similarly distributed in the two groups, and whether they showed the lessons to have had a statistically significant effect. The two sets of error rates for the pre-test and post-test are represented by boxplots in Figure 28 and mean values in Table 26. Again, the distributions seem identical, except for the outliers.[31] They are not normal, either.

Table 26. Pre-test and post-test error rates (*SOME*)

TECHNIQUE		N	Mean	Std. Deviation
corpus-based	PRE-TEST	17	1.65	1.62
	POST-TEST	17	0.71	1.05
conventional	PRE-TEST	17	1.41	1.70
	POST-TEST	17	0.59	1.06

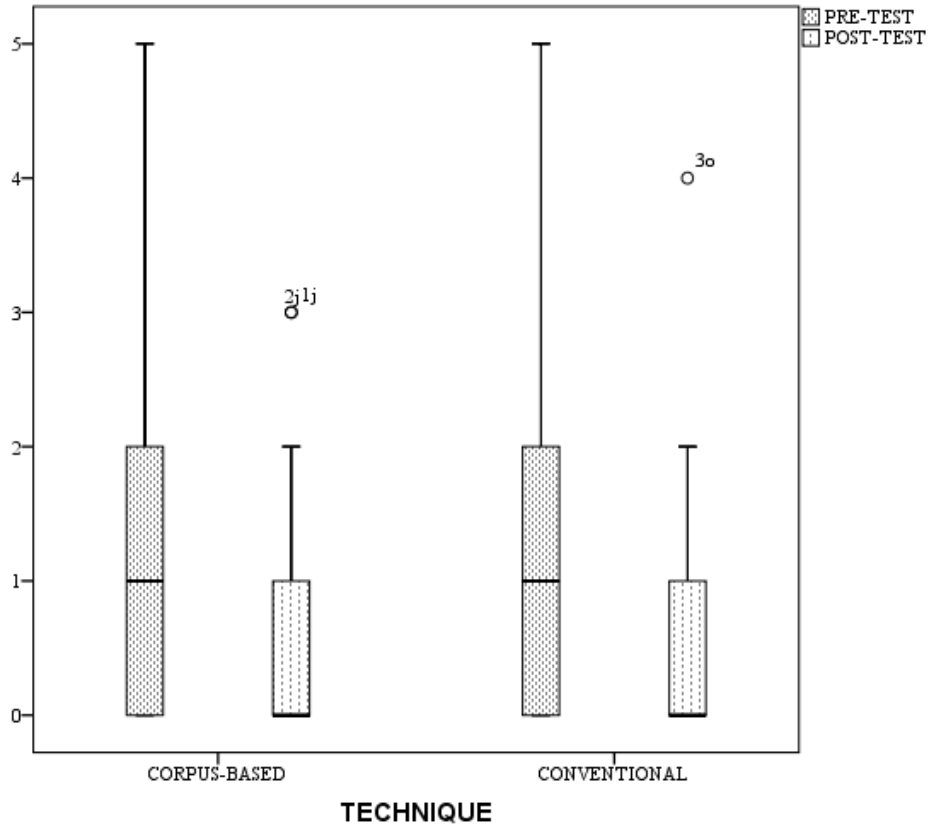


Figure 28. Boxplots for pre-test and post-test error rates (*SOME*)

[31] For an explanation of what an *outlier* is, see p. 131

Therefore, the effectiveness of the treatment was tested in a Wilcoxon Signed Ranks Test. The null hypothesis was that there would be no difference between the distributions of pre-test and post-test results in each group, which would mean that the lessons were not effective. The medians for the corpus-based group were $m_{pre} = 1$, $m_{post} = 0$, and for the conventional group they were exactly the same. The test gave the following outcome: $Z_{corp} = -2.33$, $p = .02$, $r = .56$ and $Z_{conv} = -1.91$, $p = .056$, r = .47. This means that the null hypothesis can be rejected for the corpus based group, and the alternative hypothesis, that the lesson was effective, can be accepted. The effect size for this group can be considered large. With the conventional lesson the situation is somewhat different. Although the distributions look identical on the graph, the one outlier (point *3o* on the boxplot in Figure 28) with such a high value seems to have a strong effect on the outcome of the test, so that the result ($p = .056$) falls slightly beyond the significance level accepted for the study ($\alpha = .05$). On the other hand, it is worth pointing out that the *p*-value reported here is two-tailed, which means it does not predict the direction of the effect, as this is a more conservative approach usually recommended in statistical analysis. For this particular test, however, a unidirectionality of the change may be assumed, because the lesson is expected to have a positive rather than a negative effect (a reduction in the error rate), which indeed takes place. In this case, then, the significance value obtained in the Wilcoxon test could be divided by two, which would give a result that does not exceed α ($p = .028$). The effect size of $r = .47$ can be judged as considerable. To conclude, both the corpus-based lesson and the conventional lesson can be seen as successful, and even though the change in the medians between the pre-test and post-test is not large in absolute terms, the distribution of the results has a decisively downward direction and the reduction in error rate is statistically significant. It is important to realize that unlike the gain values, the error rates cannot have negative values, so "*0*" is the lowest possible error rate that students could achieve.

Coming back to the comparison of the corpus-based and conventional lessons in terms of the gain values they generated, it must be pointed out again that the two groups' results are almost identical. The two lessons were equally successful, with the effectiveness confirmed in a Wilcoxon signed ranks test for each group. For organizational reasons it was impossible to perform a delayed test in the last stage of the study. The four delayed tests performed up to that moment were sufficient to observe that

there were no marked differences between the DDL and conventional lessons in terms of the duration of their effects.

4.10.3.6. AMOUNT

The last lesson in the experimental series was also connected with the problem of countability, or more specifically with the use of quantifiers with countable and uncountable nouns. The error in question concerns the confusion of two quantifiers: *amount of* and *number of*. The former should be used with uncountable nouns while the latter with countable nouns. Admittedly, the BNC does show *amount* used with plural nouns, as both these are words that, although plural in form, refer to such uncountable concepts as money, energy, time, and mass. At the top of the list generated by a search for plural nouns within the scope of two words to the right of *amount*, are the following: *damages*, *resources*, *calories*, *funds*, *goods*, *things*, *wages*, *times*, *shares*, and *hours*. These distinctions were included in the experimental and control lessons, the detailed instructions and materials for which can be found in Appendix D on page 430.

Table 27. General statistics for lessons on *AMOUNT*

	TECHNIQUE		Statistic	Std. Error
GAIN - *AMOUNT*	CORPUS-BASED	N	17	
		Mean	0.94	0.30
		Median	1.00	
		Variance	1.56	
		Std. Deviation	1.25	
		Minimum	-1.00	
		Maximum	4.00	
		Skewness	1.00	0.55
		Kurtosis	1.10	1.06
	CONVENTIONAL	N	17	
		Mean	1.41	0.39
		Median	1.00	
		Variance	2.63	
		Std. Deviation	1.62	
		Minimum	-2.00	
		Maximum	4.00	
		Skewness	-0.07	0.55
		Kurtosis	-0.23	1.06

An overview of data for the two lessons reveals some differences in distribution especially in skewness and kurtosis. A Shapiro-Wilk test of normality revealed that the corpus based group's gain values do not have a normal distribution. The difference is mainly in skewness, the value of which indicates a high accumulation of results in the lower range of the scale and a longer tail on the right hand side of the mean.

Overall, the distributions obtained in this experiment are less similar to each other than was the case for the previous lesson (see Figure 29).

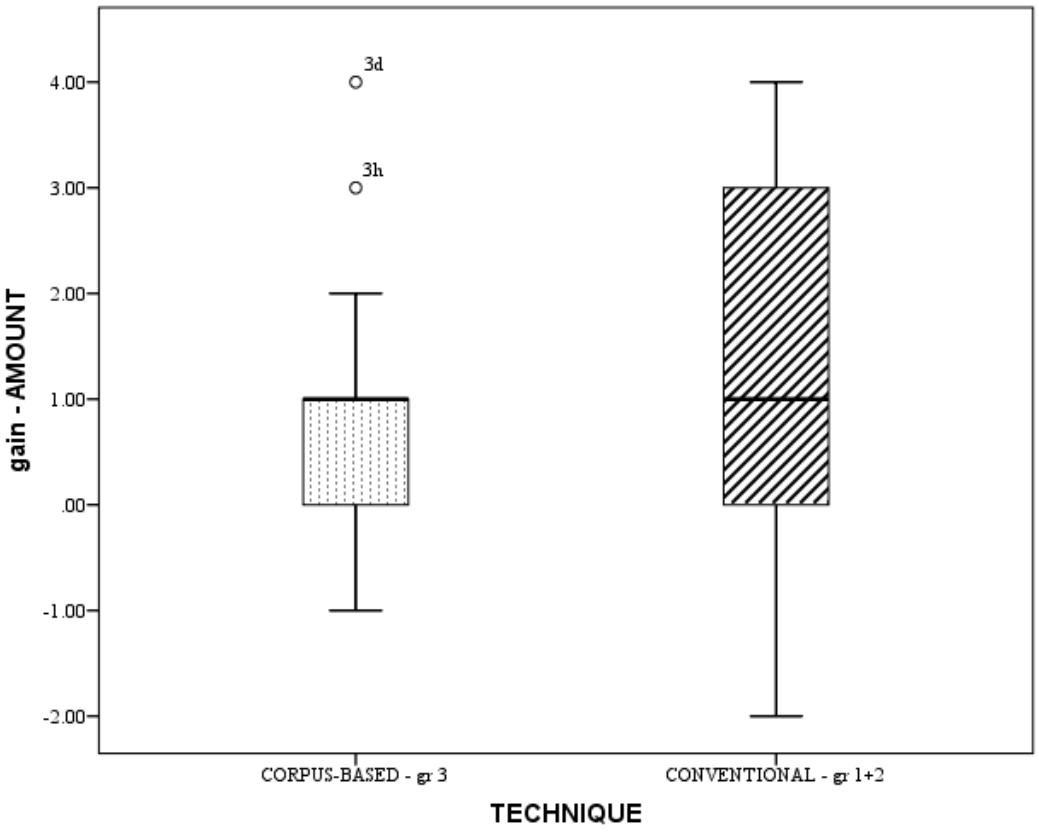


Figure 29. Boxplots for lessons on *AMOUNT* / *NUMBER* (gain values)

Though the means are different by nearly half a point, the medians are the same. The distribution for the conventionally instructed group is more balanced and more evenly spread. In the other group, most of the results are accumulated within the 0-1 band (12 out of 17), but this is probably the result of the pre-test score distribution, which was similarly compressed. This was not the case in the other group, as presented in Figure 30. After establishing a non-normal distribution of pre-test and post-test error rates in both groups, a Wilcoxon signed ranks test was performed, proving both lessons

to have been effective, as the difference in the medians for the pre-test and post-test error rates in both groups turned out to be statistically significant: $Z_{corp}(17) = -2.65$, $p = .008$, r = .64 and $Z_{conv}(17) = -2.72$, $p = .007$, r = .66. The strength of association values indicates that the effect sizes were very similar, and in fact quite large.

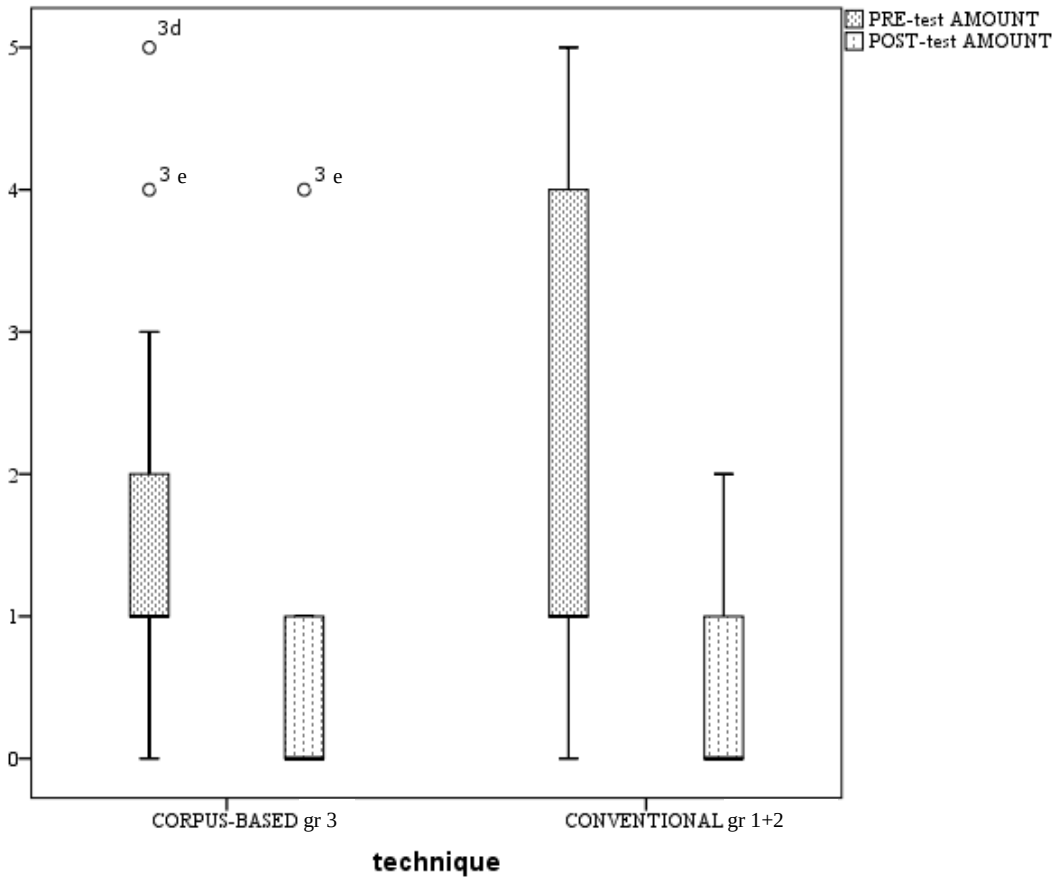


Figure 30. Boxplots for pre-test and post-test error rates (*AMOUNT*)

To conclude, the corpus-based and conventional techniques of instruction did not prove to differ in their effectiveness for the lesson on the use of *amount of* and *number of* with countable nouns. Both groups showed improvement in their error rates, with minor differences in the distribution of the scores resulting from the distribution of the error rates in the pre-test rather than from the treatment itself.

4.10.4. Overall/Combined analysis

Following the detailed analysis of the individual experimental and control lessons, a more thorough picture of the experiment is going to be presented below. All the lessons were divided into two sets, depending on which technique was used with a given group of students. And so, group 1+2 had corpus based classes on HOWEVER, STATE, and SOME, which were taught in a conventional manner to group 3. These constitute the first set. The second set holds corpus-based lessons for group 3, and the parallel conventional lessons for group 1+2. These were on POSSIBILITY, APPROVE and AMOUNT.

Students' results in these sets have been put together and analyzed collectively, so that more general conclusions could be drawn. Some more statistical tests were performed to obtain an answer to the main research question, i.e. whether corpus based teaching is more effective than alternative, more conventional teaching techniques.

4.10.4.1. Post-test raw scores analysis

One way of examining the effectiveness of the language instruction offered to students in the experimental and control lessons may be comparing the post-test error rates within each of the two sets of data specified above. At the starting point of the experiment, the distribution of the PRE-test error rates was very similar for both groups. In the first set (HOWEVER/STATE/SOME) the results in the corpus group ($M_{corp} = 1.96$, $SD_{corp} = 1.47$, $m_{corp} = 2$) were almost the same as in the control group ($M_{conv} = 2.05$, $SD_{conv} = 1.57$, $m_{conv} = 2$). In the second set (POSSIBILITY/APPROVE/AMOUNT) the difference is slightly larger ($M_{corp} = 2.06$, $SD_{corp} = 1.40$, $m_{corp} = 2$ and $M_{conv} = 2.48$, $SD_{conv} = 1.47$, $m_{conv} = 2$) but not big enough to be statistically significant, as a Mann-Whitney test indicated: $U(98) = 1001$, $p = .157$. Since the PRE-test error rates are so homogeneous, analyzing the distribution of the POST-test raw error rates apart from the gain values may well be of use. The presentation of these results will be most clear in the form of histograms, which offer precise information on the distribution of particular scores across the population sample. What might be of special interest is the number of

the most successful scores, i.e. 0s (no error for a given item) and 1s (one error). The more of these in a given set, the more effective the lessons must have been.

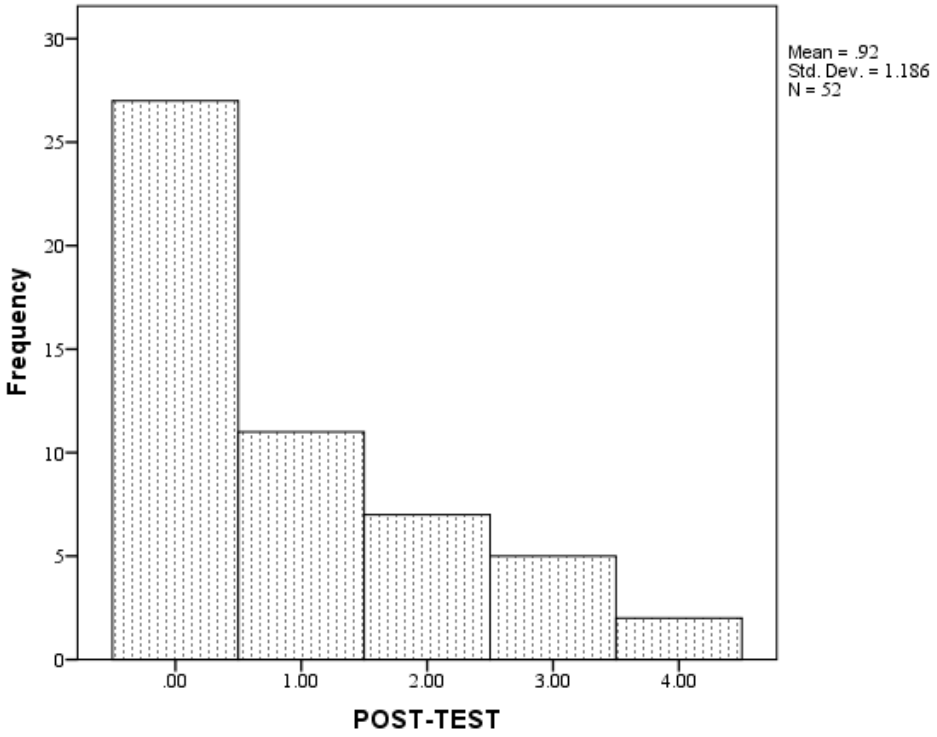


Figure 31. Error rate histogram for *HOWEVER*/*STATE*/*SOME* - CORPUS BASED (gr 1+2)

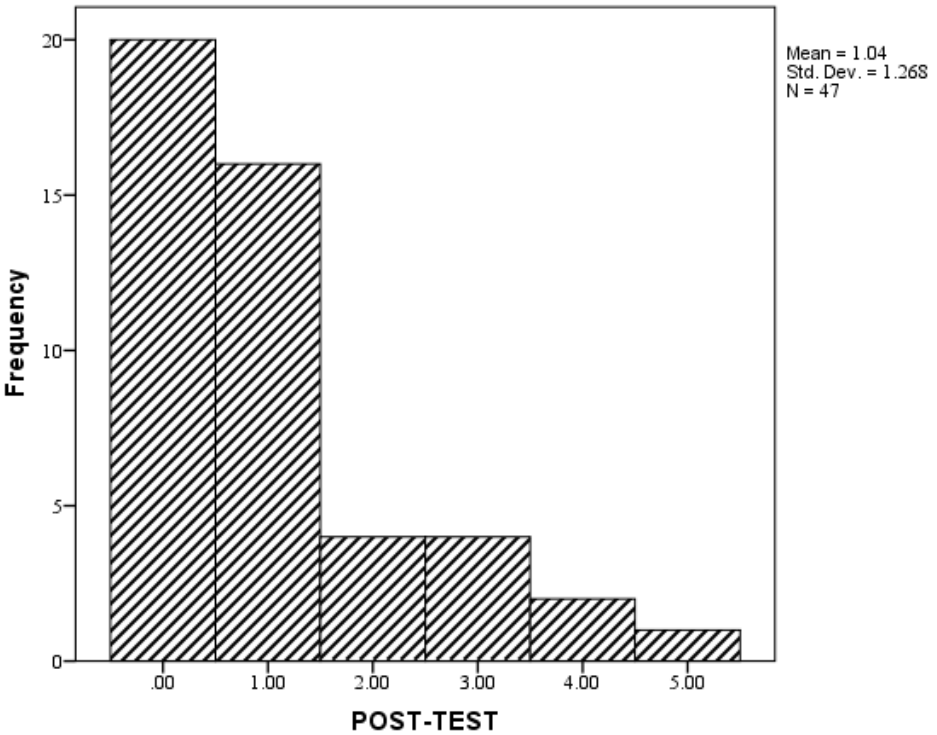


Figure 32. Error rate histogram for *HOWEVER*/*STATE*/*SOME* – CONVENTIONAL (gr 3)

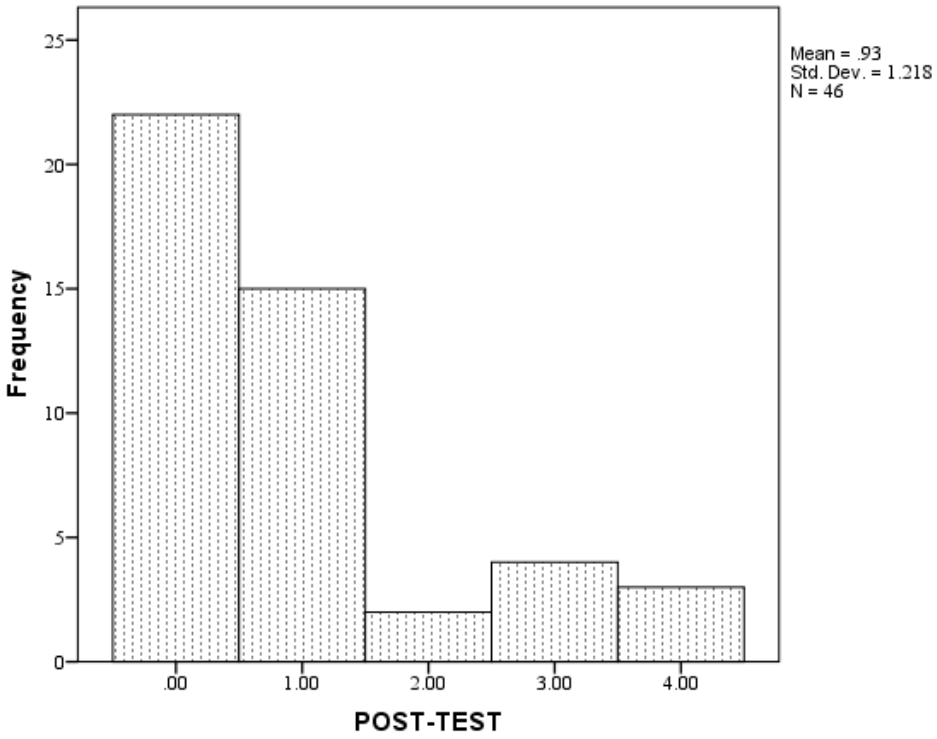


Figure 33. Error rate histogram for *POSSIBILITY/APPROVE/AMOUNT* - CORPUS-BASED (gr 3)

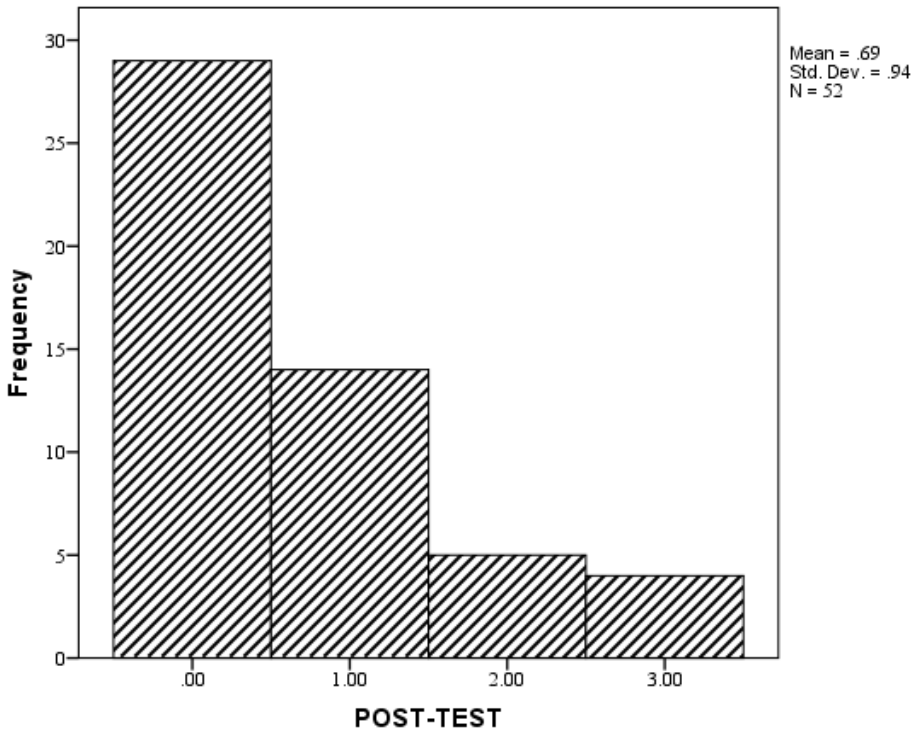


Figure 34. Error rate histograms for *POSSIBILITY/APPROVE/AMOUNT* - CONVENTIONAL (gr 1+2)

All the histograms above look very similar: in all of them the results are highly right-skewed, which is an expected and desired outcome in the post-treatment error rates: the closer to zero, the more successful the lessons have been (the fewer errors were committed). There are, however, minor differences that can be observed: after 0s and 1s (the most successful scores) were totaled for each set, it turned out that the percentage for the corpus-based groups was slightly lower. In the first set, 73 percent of the corpus-based group had 0s and 1s, as compared with 76 percent in the conventional group. In the second set, the scores were 80 percent and 82 percent respectively. Small as these percentage differences are, the trend seems to be consistent, to the disadvantage of corpus-based lessons, regardless of which student group the lessons were addressed to.

On the whole, both sets of lessons proved to be effective. A Wilcoxon Signed-Ranks test indicated that there was a significant difference between the pre-test error rates and the post-test error rates (see Table 28 for detailed test results). This confirms that students managed to improve their accuracy in the problem areas. The effect size statistics range between medium and large (Cohen 1988), so on the whole they can be considered satisfactory.

Table 28. Wilcoxon Signed-Ranks test results for pre-test/post-test error rates

ITEMS	technique	group	N	mean negative ranks	mean positive ranks	test statistic *Z*	signifi-cance *p* (2-tailed)	effect size *r*
HOWEVER/ STATE/ SOME	corpus-based	1+2	47	20.58	14.71	-3.94	< 0.001	0.41
	conventional	3	52	18.37	14.14	-3.43	0.001	0.34
POSSIBILITY/ APPROVE/ AMOUNT	corpus-based	3	52	17.43	11.50	-3.93	< 0.001	0.39
	conventional	1+2	46	23.22	10.13	-5.26	< 0.001	0.55

The conventional lessons with group 1+2 in the second set stand out as particularly effective. The high effect size value is a reflection of the highest mean of negative ranks (which occurred when there was a lower number of errors in the post-test than in the pre-test) compared with the other sets, and the lowest mean of the positive ranks (where a student had fewer errors in the pre-test than in the post-test). What is more, this group of lessons recorded the lowest number of ties: 9, as opposed to 13-14 in the other three sets. Ties are unchanged values between the two takes, and so signify no

progress in the learner's control of the problem, so the fewer of them, the better. The same group of students had a higher effect size value in the other set of lessons as well (r = .41), which may indicate that this was generally a more 'teachable' group than the other one, regardless of the technique. This corresponds with the instructor's subjective perception. The difference between this group's results in the two sets of lessons could be attributed to the students' preference for conventional styles of teaching. The survey results, discussed in section 4.11. may help resolve this issue.

4.10.4.2. Combined analysis of gain values

The lesson-by-lesson analyses, presented in section 4.10.3. were based mainly on gain values. It may be useful to see the same data gathered together and try to make some generalizations about them. Figure 35 and Figure 36 below present the distributions of the gain values for all lessons in the experiment. The graphs for the conventional lessons tend to have more symmetrical distributions, with medians that are closer in values than is the case for the corpus-based lessons.

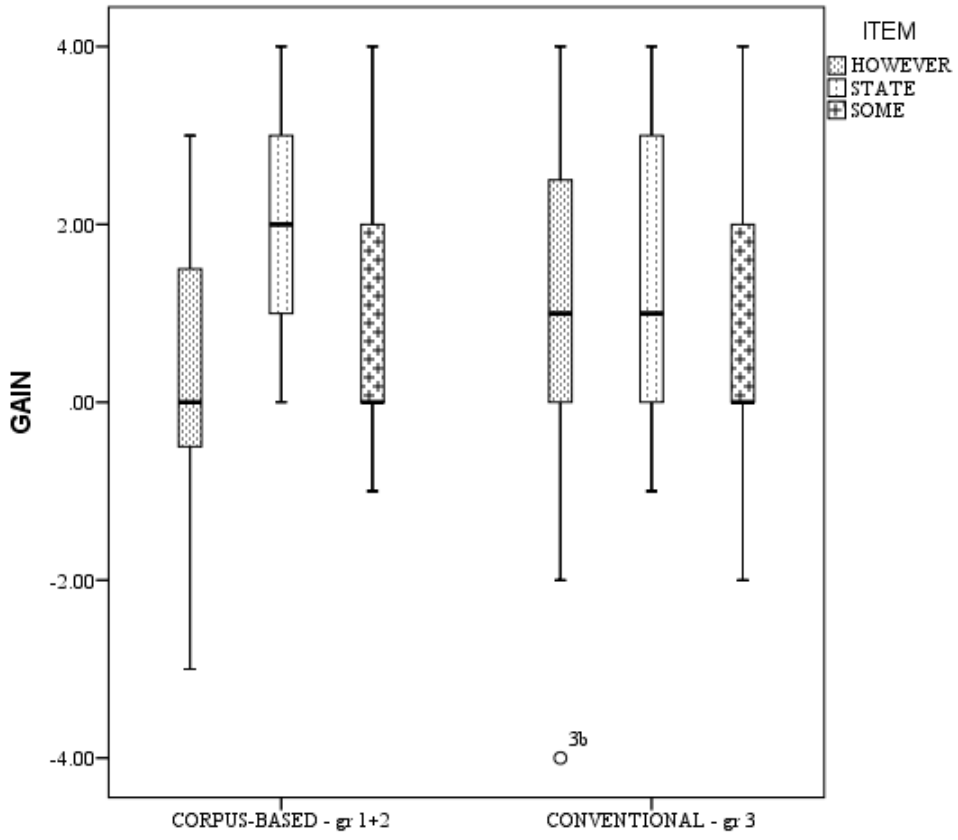


Figure 35. Comparison of boxplots for *HOWEVER*, *STATE* and *SOME*

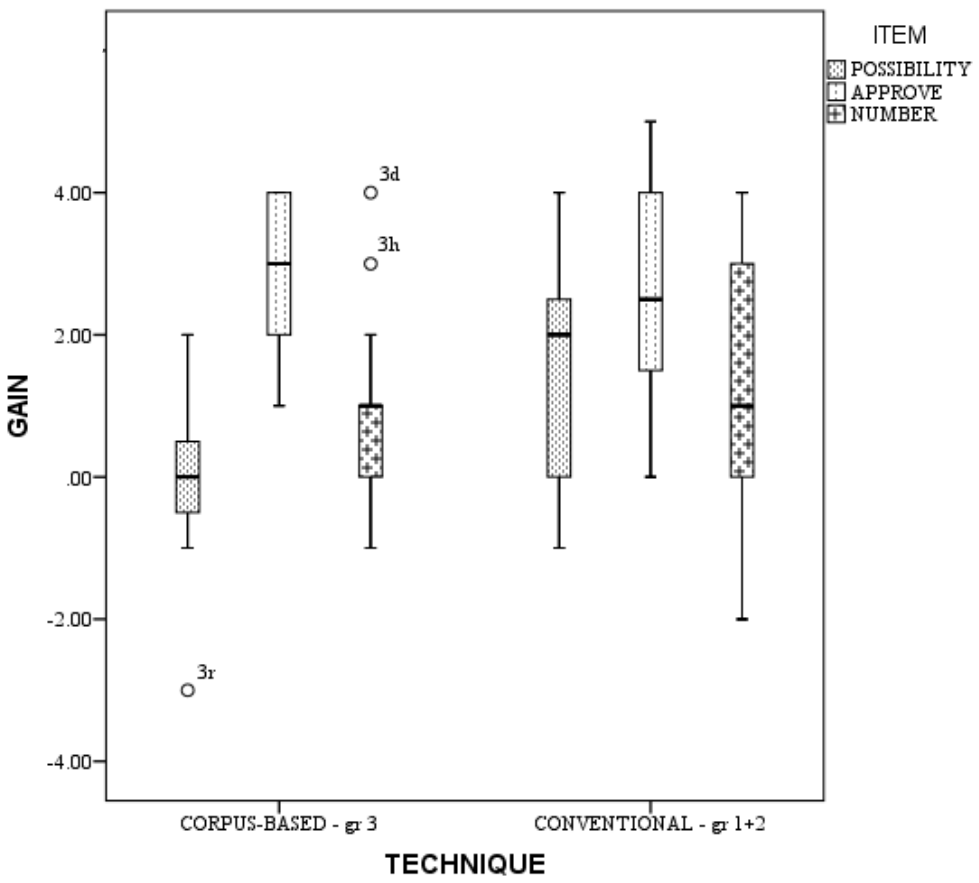


Figure 36. Comparison of boxplots for POSSIBILITY, APPROVE and AMOUNT

This impression was verified by the independent samples Kruskal-Wallis H test, a non-parametric equivalent of the one-way ANOVA, which allows the comparison of more than two independent groups. Even though the lessons were performed with the same students, every time the composition of the group was slightly different and the topics of the lessons differed, so those lessons were assumed to be independent observations performed in different conditions; therefore, the test was found to be applicable to the situation of the study. The aim of the test was to establish whether the distributions of gain results for the conventional lessons in each set were consistent with one another, and then the same was done for the corpus-based lessons. In both sets of lessons, for the *conventional* component the test did not reject the null hypothesis that the distribution of gain values is the same across the categories of ITEM, i.e. for the three lesson topics. For the first set (*however/state/some*), $H(2) = 0.79$, $p = .67$ (two-sided), with a mean rank of 24.1 for HOWEVER, 26.3 for STATE, and 22.0 for SOME. In the second set,

the situation is similar, though the results are much closer to the threshold value of $\alpha = .05$ assumed for the whole study: $H(2) = 4.90$, $p = .086$, with a mean rank of 23.5 for POSSIBILITY, 33.4 for APPROVE, and 23.4 for AMOUNT. This indicates that no statistical difference was found between the gain values for the conventional lessons in each set. This outcome could be interpreted to mean that effectiveness of conventional lessons is quite even and is unaffected by their subject matter.

As far as corpus-based lessons are concerned, the Kruskal-Wallis H test results did reveal significant differences between the gains for different lessons. In the first set (*however/state/some*), the test statistic was $H(2) = 9.71$, $p = .008$, and the mean ranks are as follows: 20.5 (HOWEVER), 35.8 (STATE) and 24.4 (SOME). Pairwise comparisons, which are performed if the test result is significant, revealed that the distribution of results for HOWEVER is significantly different from that for STATE ($p = .007$). The other pairs of items did not prove to be significantly different. In the second set of data (*possibility/approve/amount*) the differences are still more prominent: $H(2) = 22.46$, $p = <.001$, with the mean ranks of 13.8 (POSSIBILITY), 36.9 (APPROVE), and 22.4 (AMOUNT). Pairwise comparisons were made again, and showed significant differences in the distribution between two pairs of items: POSSIBILITY and APPROVE ($p < .001$) and APPROVE and AMOUNT ($p = .008$). The only pair that did not prove to be significantly different was AMOUNT and POSSIBILITY ($p = .185$).

The results of the above tests indicate that the effectiveness of corpus-based lessons is more easily affected by whatever various factors may be involved, for example the choice of the subject matter that they address. While some language problems will be well suited to the corpus-based lesson design, many others may not really give the expected results. It seems that the most appropriate material for this type of lessons would be that of a clearly *lexical* nature, with a limited set of options in terms of word choice or complementation. Where lessons involve more abstract categories like, for example, countability, or lexicogrammatical issues that need to be considered in a broader context, corpus-based lessons give less satisfactory results. Another explanation for the results obtained in the Kruskal-Wallis H tests for the DDL lessons may be that the lessons on HOWEVER and POSSIBILITY were the first in the series and so may have confused the students slightly. Only with time did the corpus-based techniques become more successful. The survey results seem to confirm such an interpretation. It

may be that both of these factors (type of language problem and novelty) contributed to the final outcome in some measure.

The distribution of gain values obtained in the experiment indicates that the effectiveness of corpus-based teaching is less predictable and less consistent. It may be more easily affected not only by what material is being taught, but also by various learner factors, such as cognitive styles or attitudes to this innovative teaching technique itself. There is an indication in the results presented above that the corpus-based teaching techniques might not have satisfied the high expectations the researcher had had at the beginning of the project, especially in terms of effectiveness. The good results with typically lexical items must be noted, however.

4.10.4.3. Analysis of pooled gain values

The broadest view of the data should be obtained after they have been pooled together within the previously defined sets, so that more general comparisons between corpus-based and conventional lessons can be made. In this way the relatively small number of participants can be compensated for. At the same time, the fact that a series of lessons on a variety of language problems was administered should render a more comprehensive picture of the outcome of the experiment. Figure 37 presents the gain values for both the corpus-based and the conventional lessons on HOWEVER, STATE and SOME.

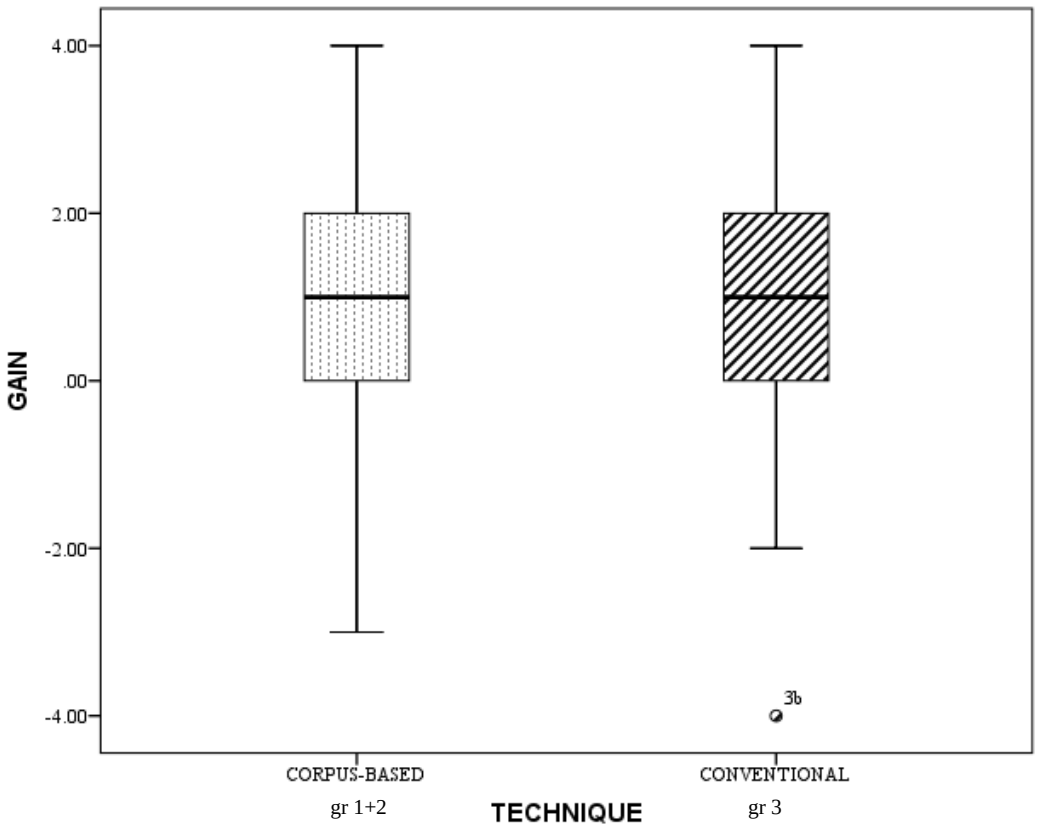


Figure 37. Pooled gain values for *HOWEVER*, *STATE* and *SOME*

As can be seen from the graph, the distributions of the two sub-sets are almost identical. There is a minor difference in the lower sections of the distributions, but on the whole the similarities are evident: with the same median and the same position of the upper and lower hinges of the boxes. Some information on the distribution must be supplemented, as it is not provided in the graph: the mean and standard deviation values are $M_{corp} = 1.04$, $SD_{corp} = 1.6$ and $M_{conv} = 1.00$, $SD_{conv} = 1.76$. No statistical test is needed to confirm the similarity of the distributions, as both the medians and means are the same, and the standard deviations are very much alike, as well.

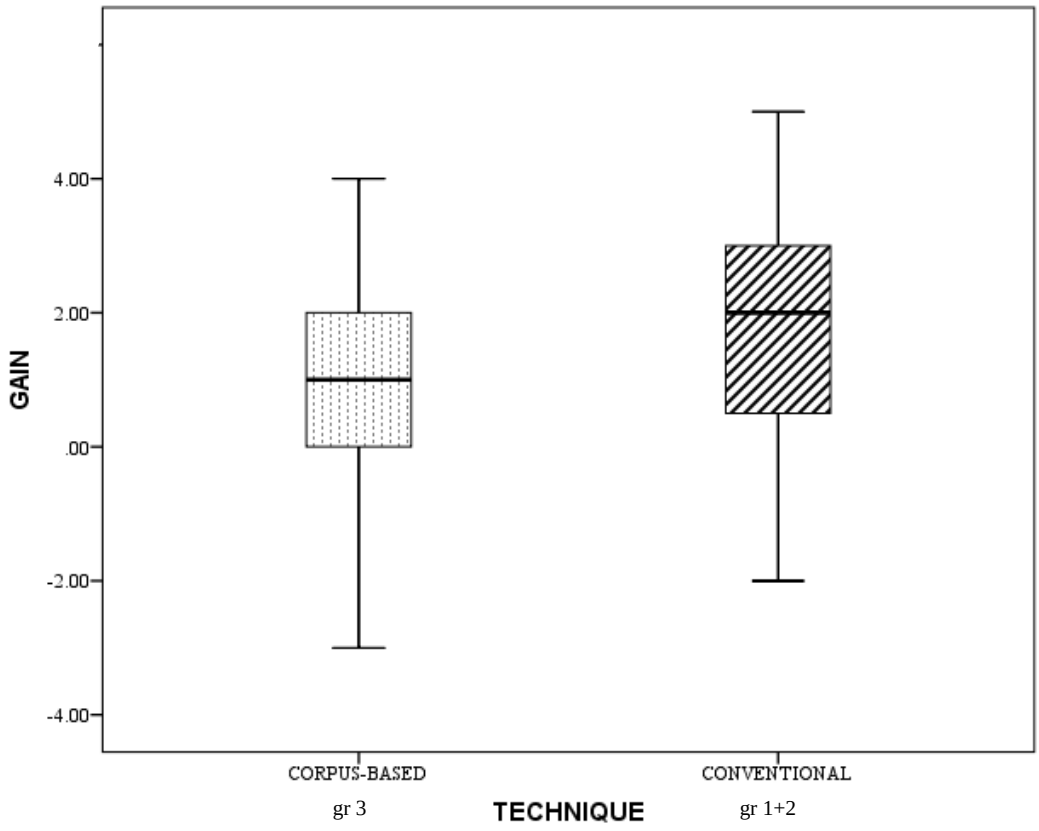


Figure 38. Pooled gain values for *POSSIBILITY*, *APPROVE* and *AMOUNT*

The situation in the other set (POSSIBILITY, APPROVE and AMOUNT) is less clear, though. The shapes of the distributions presented in Figure 38 are similar again, but the values are not: $M_{corp} = 1.13$, $SD_{corp} = 1.65$, $m_{corp} = 1$ and $M_{conv} = 1.79$, $SD_{conv} = 1.70$, $m_{conv} = 2$. The difference seemed worth examining, and so another statistical test needed to be performed. The test whose assumptions correspond most closely to the situation in the experiment is the Mann-Whitney independent samples test. Admittedly, the samples *within* each set are not totally independent as they come from the same group of students tested on three occasions, but the collections of samples that are being compared in the test are independent of each other. The dependent variable, that is the factor that is being tested, is the same in each case: the effect of the technique of instruction on the error rate gain value.

Therefore, a Mann-Whitney test was performed, and its results are somewhat ambiguous: the test outcome is on the verge of the accepted significance level ($U = 927.5$, $p = .052$, $r = .20$). The assumed null hypothesis that there is no difference in

distribution between the gain values for the corpus-based and conventional lessons in the second set will not, therefore, be rejected ($p > \alpha$), but the effect size (r) indicates that the difference is noticeable, though not very large. Therefore, it must be said that the test did not show the difference in effectiveness between the corpus-based and conventional lessons to be statistically significant at the confidence level of 95%; the result is, however, marginal and there is only a 5.2% likelihood that the lower gain scores for the corpus-based classes on POSSIBILITY, APPROVE and AMOUNT could be due to chance, after all. Whether this was actually the case or not should remain an open question, and further studies on larger groups of students would be needed to resolve the issue. The differences evidenced within the same set of lessons in the discussion above indicate that some of them were, indeed, a disappointment.

4.11. The survey

The experiment reported above has been supplemented with another source of information on corpus-based lessons – a survey that was administered to its participants. Evidence of the effectiveness of an innovative educational procedure is not enough by itself to deem it successful. It also needs to be accepted by the learners. The purpose of the survey was to find out what students' attitudes to corpus-based teaching were, as compared with some other well established classroom activities. Apart from these attitudes, the survey was planned to ask students about their convictions regarding degrees of effectiveness of various teaching techniques, as well as examine interdependencies between those opinions and students' actual achievements. Considering that the subjects of the study were just graduating from a teacher training college and some of them would eventually become teachers themselves, the survey results may be of more general interest to teacher trainers. The survey was carried out in Polish, the students' L1, so that any misunderstandings and comprehension problems could be avoided. Its full text has been placed in Appendix F, p. 37, together with survey data not included in the main body of the chapter.

4.11.1. Background information

The first part of the survey collected information about the 35 respondents: their name, and hence gender (identification was necessary for correlational analysis involving their achievements in English), educational status and prospects, their self-assessed level of proficiency in English, their L3s and their level of proficiency in them.

As is the case in most philology departments in Poland, a large majority of students in the group were female (27 out of 35). Eight students had either completed or were in the course of other studies. Almost all of them (31) declared their willingness to continue their education at the MA level. When asked to assess their proficiency in English as compared with the expectations they had had before they began their studies in KJO, 24 rated themselves as GOOD, 6 as AVERAGE, 4 as VERY GOOD, and 1 as POOR. As for foreign languages that they had started to learn before English, the most popular was German (6), but generally there were very few positive responses in this category. Over half of the students declared to know one foreign language other than English (18), 16 declared two languages and one student three. These numbers include six students who decided to list Latin, which all of them had as an obligatory course at KJO. Others did not list it among the foreign languages they knew. The level of proficiency in L3s declared by the students was spread quite evenly between *elementary* (13), *lower-intermediate* (11), and *intermediate* (11). When there was more than one L3 declared, the highest level of proficiency among them was registered for this index. The intention behind the questions on L3s was to obtain information about individual students' language aptitude and interest in language in general. A possibility was considered that these data could correlate with how various students perceived corpus-based lessons and how they responded to them.

4.11.2. Item 1

In response to the first item of the questionnaire proper, the respondents were to decide which of the ten adjectives listed for them expressed their opinions on corpus-based activities. Five of the adjectives had a positive bias, and five – negative. The sequence

of the phrases was randomized in the questionnaire, but Table 29 and Table 30 present them as sorted into positive and negative groups for the sake of clarity:

Table 29. Positive adjectives – "YES" responses

		Group		Total N=35	Percent of population
		group 1+2 N = 18	group 3 N = 17		
Positive adjectives	convincing	12	15	27	77.1%
	interesting	12	12	24	68.6%
	effective	11	12	23	65.7%
	thought-provoking	16	14	30	85.7%
	clear and understandable	12	14	26	74.3%

Table 30. Negative adjectives – "YES" responses

		Group		Total N=35	Percent of population
		group 1+2 N = 18	group 3 N = 17		
Negative adjectives	boring	2	2	4	11.4%
	confusing	3	0	3	8.6%
	too difficult	0	1	1	2.9%
	overwhelming	3	4	7	20.0%
	too time-consuming	4	3	7	20.0%

The responses indicate that the students' attitude to corpus-based teaching was generally positive, although the item's main task was to lead the respondents into the topic and make them start analyzing their position on the issue. Perhaps it is worth mentioning that "no" was given in response to positive adjectives only 14 times, compared with the 130 "yeses" listed in Table 29 above. As for the negative adjectives, the proportion is reversed: 22 to 133. "Don't know" or no answer was given 31 times for the positive adjectives and 20 times for the negative ones. Table 31 shows the per capita values of the number of times the adjectives were given a "yes" in each sub-set. There is a slight difference between the two groups' means, but it is not statistically significant. (Since the distribution is not normal, a non-parametric test has been applied to verify the similarity of distributions between the two groups.):

Table 31. Mean results for positive and negative adjectives

Group		Positive adjectives	Negative adjectives
group 1+2	N	18	18
	Mean	3.44	0.67
	SD	1.76	0.91
	Median	4	0
group 3	N	17	17
	Mean	3.94	0.59
	SD	1.09	0.87
	Median	4	0
Total	N	35	35
	Mean	3.69	0.63
	SD	1.47	0.88
	Median	4	0
Mann-Whitney U test (groups)	*U*	140	144
	Z	-0.45	-0.34
	p	0.65	0.74

In some cases the differences between groups 1+2 and 3 are minimal, and this is where results will be presented in total only. Wherever relevant, the data will be split into groups and then given in total.

As stated above, the students' attitudes to DDL expressed in responses to item 1 are mostly positive. The highest number of students chose the adjective *thought-provoking* (30), and out of the positive adjectives the least common one was *effective*, the feature that is crucial to this study. The items that follow should give a better insight into the students' opinions on the matter.

4.11.3. Item 2

Here students were asked to list language problems discussed in corpus-based lessons carried out in the grammar course. This was supposed to verify their answers to one of the items in the next set, when they were asked whether corpus-based classes helped them remember the language problems discussed. Admittedly, the fact that students can or cannot remember in what way they learned particular material does not need to be connected with how well they actually learned it. The general idea behind the item was to see how memorable these lessons were to students. Table 32 presents the distribution of their answers to Item 2.

Table 32. Responses to item 2

		Group		Total	Valid percent
		group 1+2	group 3		
No. of language problems remembered correctly as taught in a corpus lesson	0	1	3	4	11.1
	1	7	3	10	28.6
	2	6	8	14	40.0
	3	4	3	7	20.0
Total		18	17	35	100.0

Considering that the first of the classes took place six months before the questionnaire was administered, and that not all students were present at all the classes, the answers are moderately satisfying.

4.11.4. Item 3

The next item required the students to determine their position on three statements regarding corpus-based lessons:

(1) Tasks based on corpus data help me *understand* language problems better than other types of activities;
(2) Tasks based on corpus data help me *remember* language problems better than other activities;
(3) Tasks based on corpus data are more *effective* than other activities, i.e. they help me achieve better accuracy in the language areas addressed by the lesson.

Responses to these statements were constructed on a 5-point Likert scale from "strongly disagree" (1), through "disagree" (2), "no opinion" (3), and "agree" (4), to "strongly agree" (5). In order to verify the reliability of the survey item, Cronbach's Alpha was calculated for it ($\alpha = .77$). The following results were collected:

Table 33. Responses to item 3

		Group		Total	Valid percent
		group 1+2	group 3		
understand	strongly disagree	2	0	2	5.7%
	disagree	3	2	5	14.3%
	no opinion	3	5	8	22.9%
	agree	9	8	17	48.6%
	strongly agree	1	2	3	8.6%
remember	strongly disagree	2	0	2	5.7%
	disagree	4	2	6	17.1%
	no opinion	3	1	4	11.4%
	agree	7	13	20	57.1%
	strongly agree	2	1	3	8.6%
effective	strongly disagree	1	0	1	2.9%
	disagree	4	3	7	20.0%
	no opinion	10	7	17	48.6%
	agree	3	6	9	25.7%
	strongly agree	0	1	1	2.9%

As Table 33 shows, "effective" was the word most of the respondents did not feel like choosing in reference to the DDL techniques, despite having been given a clear explanation of what exactly effectiveness is meant to be here. This confirms the initial observations made above in reference to responses to item 1: Students are not confident of the outcome of these lessons, although many still seem to have an open mind about them. A majority of them believe that the DDL lessons help them *remember* (65.7%) and *understand* (57.2%) the material, but fewer students are equally confident that their English actually improves as a result (only 28.6%). The negative answers (*strongly disagree* and *disagree*) amount to around 20% in each case, which probably corresponds to the proportion of students who are not keen on corpus-based techniques in general. This proportion is not negligible.

Table 34. Mean results for item 3

Group		understand	remember	effective
Group 1+2	N	18	18	18
	Mean	3.22	3.17	2.83
	Std. Deviation	1.17	1.25	0.79
	Median	4.00	3.50	3.00
group 3	N	17	17	17
	Mean	3.59	3.76	3.29
	Std. Deviation	0.87	0.75	0.85
	Median	4.00	4.00	3.00
Total	N	35	35	35
	Mean	3.40	3.46	3.06
	Std. Deviation	1.04	1.07	0.84
	Median	4.00	4.00	3.00

Table 34 above shows the same information in a more compact way, which permits an additional observation: group 1+2 shows lower mean values across all the three parts of this item, and the medians are the lowest for it in both groups. Their attitude seems on the whole more skeptical or cautious, especially in response to the item on the *effectiveness* of DDL. This seems to indicate that indeed their particularly high gains in the conventional set of lessons might be connected with their preference for traditional instructional techniques.

4.11.5. Item 4

Item four was worded in the questionnaire in the following way (translated into English): "How do you assess the effectiveness of the types of classroom activities listed below, in cases when the subject matter of the lesson is a feature particularly difficult to Polish learners of English? Base your answer on your own experience."

For the sake of analysis the activities have been divided into three groups, depending on what type of context was provided for the target language form: corpus-based, sentence-based, and text-based. Students' responses are presented in the form of mean results for every activity they evaluated, and then as pooled together for each of the three sets. Since some differences between the student groups have emerged, the results are presented for the groups and in total.

Table 35. Mean results for individual corpus-based techniques

Group		NS corpus conc.	parallel corpus	partial transl	EC - conc	gapped conc.	frequency lists	OVERALL MEAN
group 1+2	Mean	3.50	3.33	3.53	3.43	3.56	3.06	3.42
	N	16	18	17	14	16	16	
	SD	0.52	0.77	0.62	0.76	0.81	1.00	
	Median	3.50	3.00	4.00	4.00	4.00	3.00	
group 3	Mean	3.81	3.88	3.73	3.88	3.80	3.31	3.74
	N	16	16	15	16	15	16	
	SD	0.66	0.81	0.80	0.81	0.68	0.87	
	Median	4.00	4.00	4.00	4.00	4.00	3.00	
Total	Mean	3.66	3.59	3.63	3.67	3.68	3.19	3.58
	N	32	34	32	30	31	32	
	SD	0.60	0.82	0.71	0.80	0.75	0.93	
	Median	4.00	4.00	4.00	4.00	4.00	3.00	

Out of the corpus-based techniques placed in the item, *frequency lists* seems to have the weakest support, as the mean and median both indicate "average" to have been the most popular choice. Only 13 students answered "high" or "very high" here, which is the lowest result in this category, and second lowest altogether – after *word maps* (9), which belong to none of the three sets and have been added to the list so that some feedback could be obtained about another innovative technique introduced in the course. Otherwise there does not appear to be much variety in the students' preferences between corpus-based tasks, except that group 1+2 again shows consistently lower values in the evaluation. An independent samples t-test was performed to compare the two groups' effectiveness judgments, as expressed by the mean value on the Likert scale for the corpus-based techniques. There was a significant difference in the results between group 1+2 ($M_{1+2} = 3.42$, $SD_{1+2} = 0.41$) and group 3 ($M_3 = 3.74$, $SD_3 = 0.47$) for all the corpus-based activities combined: $t(33) = -2.12$, $p = .042$ (2-tailed). The result indicates that indeed the two groups evaluate corpus-based techniques differently, group 1+2 showing more reticence. What is also worth noting, no significant differences between the groups were found for their evaluation scores of either sentence-based or text-based activities. Table 36 accumulates all the Likert scale responses for the six DDL techniques listed in the questionnaire, thus showing the differences more directly:

Table 36. Pooled responses for corpus-based techniques

		Group		Total
		group 1+2	group 3	
Effectiveness of corpus-based techniques	none	1	0	1
	limited	11	5	16
	average	35	29	64
	high	48	46	94
	very high	2	14	16
N		18	17	35

The main discrepancy seems to be in the high number of enthusiasts in group 3, who gave the highest score, and the very low number of these in group 1+2. Also, there are more than twice as many negative 'votes' in group 1+2. On the whole, however, positive or neutral responses prevail, so it cannot be said that either of the groups is reluctant or outrightly hostile to corpus-based teaching; one is merely less enthusiastic than the other. The percentages of responses to each option on the Likert scale for DDL techniques are presented in Figure 39.

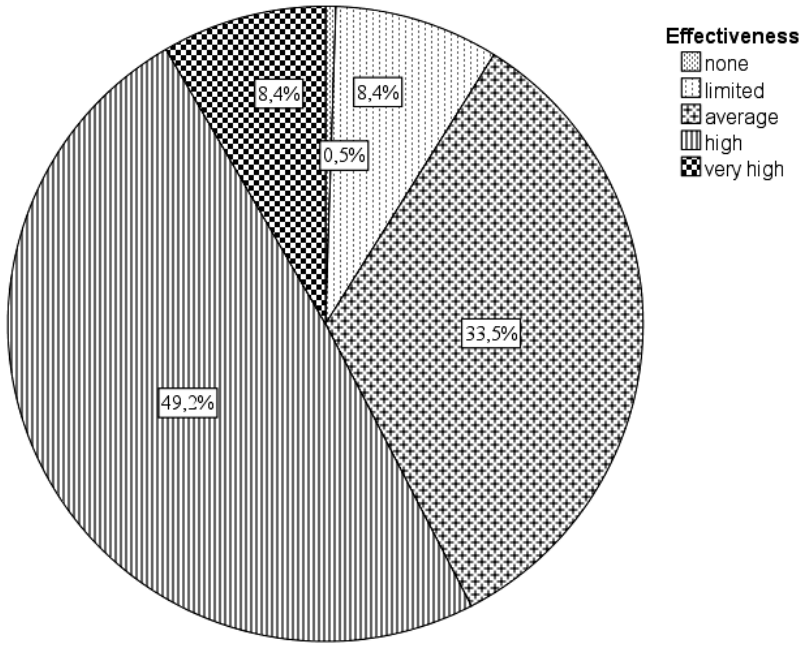


Figure 39. Pie chart for students' evaluation of corpus-based techniques

The sentence-based techniques were much less controversial: the groups were quite consonant about the techniques listed and these were evaluated very evenly. Perhaps the only exception was the enthusiastic approval of transformations (14 "very

high" votes – the highest score of all). Admittedly, this very traditional and rather automatic type of exercise did not feature very often in the third-year grammar course the students were just completing. It may be that they felt comfortable with the task, having known it and used it for many years, and found it perhaps somehow reassuring. Its infrequent occurrence in class may have been a disappointment to the students in both groups, hence its enhanced score in the survey.

Table 37. Mean results for individual sentence-based techniques

Group		teacher's examples	translation - sentences	EC - sentences	synthesis	transfor-mations	sentence building	OVERALL MEAN
group 1+2	Mean	3.83	3.56	3.65	3.61	4.11	4.06	3.80
	N	18	18	17	18	18	18	
	SD	0.79	0.92	0.49	0.85	0.83	0.87	
	Median	4.00	4.00	4.00	4.00	4.00	4.00	
group 3	Mean	3.94	3.53	3.53	3.50	4.18	3.94	3.78
	N	17	17	15	16	17	17	
	SD	0.43	0.94	1.19	0.89	0.88	0.66	
	Median	4.00	4.00	3.00	3.50	4.00	4.00	
Total	Mean	3.89	3.54	3.59	3.56	4.14	4.00	3.79
	N	35	35	32	34	35	35	
	SD	0.63	0.92	0.88	0.86	0.85	0.77	
	Median	4.00	4.00	4.00	4.00	4.00	4.00	

It is worth pointing out a difference between group 1+2 and group 3 in their evaluation of the one sentence-based task, i.e. error correction (see Table 37). Although the mean values are not very far apart, differences in the standard deviation indicate that the results are distributed very differently. Indeed, group 1+2 seems relatively uniform in its members' opinions: 11 students selected "high" and 6 "average". No other responses were given. In the other group the whole range of responses was employed, with the following scores from lowest to highest: 1, 1, 6, 3, and 4. Such a distribution gives the unusually high value of the SD, much higher than for any other task in this set, but a similar mean value to the one in the other group. Note that the median here is 3 (corresponding to *average*) while in the other group it is 4, (*high*). This discrepancy may be related to the ambiguous feelings many students usually express in reference to the error-correction task: they find it extremely difficult, but at the same time they are aware that, as future teachers, they need to develop the skills and competencies the task requires. It is, however, difficult to explain why such a variety of opinion appeared in one group and not in the other.

Table 38. Pooled responses for sentence-based techniques

		Group		Total
		group 1+2	group 3	
Effectiveness of sentence-based techniques	none	0	2	2
	limited	8	4	12
	average	24	27	51
	high	56	47	103
	very high	19	19	38
N		18	17	35

The accumulation of all the responses regarding the sentence-based activities (Table 38 and Figure 40) shows that students felt more confident about them than they did about corpus-based tasks. Neutral and negative opinions constitute less than 30% of the responses, which indicates that students felt very confident about this group of classroom activities and their effectiveness. Clearly, there are advantages to sentence-based materials: they are concise, usually short and easy to understand. The problem is that the transition from such artificially created material to real-life communication may be more difficult, as learners are not provided with any clues as to a possible situation in which a given sentence could be produced, or to the intention behind the message.

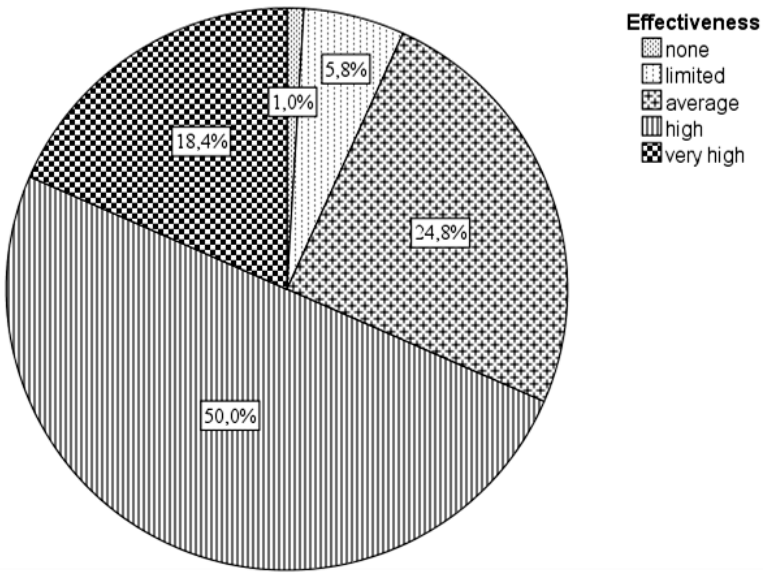


Figure 40. Pie chart for students' evaluation of sentence-based techniques

The last group of activities placed in item 4 is text-based tasks. In these activities target forms appear in context, so one could expect that preferences for these and for corpus-based activities would be parallel. To some extent this is the case: the group that was less enthusiastic about the latter has lower values here as well, except that the

difference is smaller (about 0.10 – 0.15% for each task). This is not much, but considering that the trend is consistent, it does indicate that those students had a preference for working on isolated sentences rather than authentic language material.

Table 39.Mean results for individual text-based techniques

Group		translation - fragments	EC - text	text building	OVERALL MEAN
group 1+2	Mean	3.65	3.61	3.65	3.67
	N	17	18	17	
	SD	1.00	0.78	1.12	
	Median	4.00	4.00	4.00	
group 3	Mean	3.76	3.75	3.81	3.78
	N	17	16	16	
	SD	0.83	1.00	0.75	
	Median	4.00	4.00	4.00	
Total	Mean	3.71	3.68	3.73	3.72
	N	34	34	33	
	SD	0.91	0.88	0.94	
	Median	4.00	4.00	4.00	

Table 40. Pooled responses for text-based techniques

		Group		Total
		group 1+2	group 3	
Effectiveness of text-based techniques	none	1	1	2
	limited	4	1	5
	average	18	15	33
	high	19	23	42
	very high	10	9	19
N		18	17	35

The pie chart for the text-based activities also looks similar to that for the corpus-based set, except that the proportions between the two highest scores have changed, so that there are more counts of *very high* and fewer of *high*. Again, as in the case of the transformations, familiarity may be a factor. Students may be more willing to grant the highest scores to tasks that they know very well and have a lot of experience with.

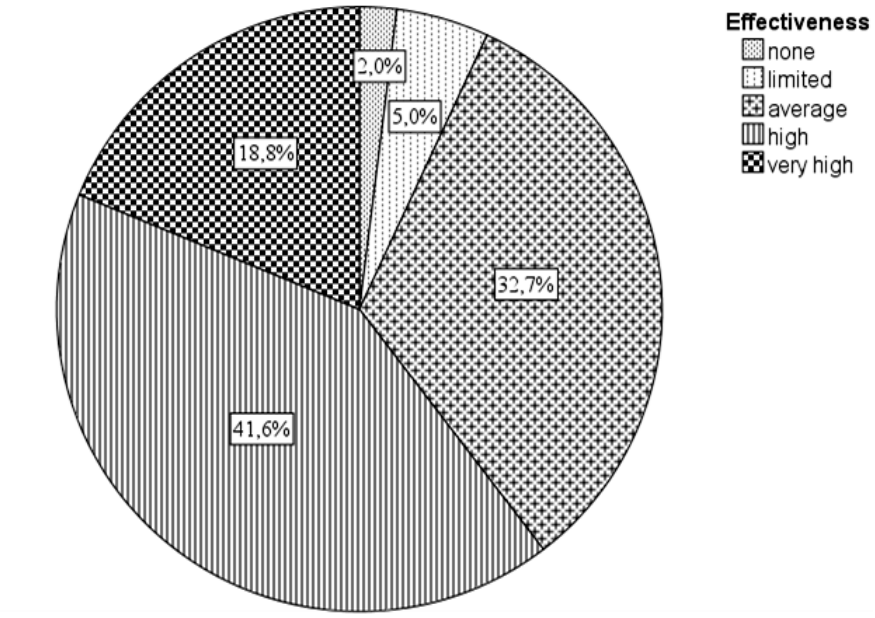


Figure 41. Pie chart for students' evaluation of text-based techniques

Tables with results for each technique include a mean Likert scale value for each student group separately and in total. A comparison of the total values reveals that they are the lowest for the corpus-based activities (M_{corp} = 3.58) while the other two sets are very close to each other: M_{sent} = 3.79 and M_{text} = 3.72). Three dependent samples t-tests were performed to compare the differences between all pairs in this triangle of variables, and the difference between the mean assessment values for the corpus-based activities and the sentence-based activities was identified as significant: $t(34) = -2.7$, $p = .010$ (2-tailed). This indicates that the lower mean values for DDL lessons are not coincidental and they reflect the fact that corpus-based activities are perceived as less effective than sentence-based ones. The other contrasts did not prove significant.

Item 4 offered some insight into how students perceived the effectiveness of various classroom activities, especially in those areas of English which cause most problems to Polish learners. It must be said that the mean value obtained for the corpus set was the lowest: $M_{corp} = 3.58$, $SD_{corp} = .37$, as compared with sentence-based tasks ($M_{sent} = 3.79$, $SD_{sent} = .40$) and text-based tasks ($M_{text} = 3.72$, $SD_{text} = .68$). A dependent samples t-test was performed for each pair of variables in this triangle, and the results showed that students assessed the effectiveness of the corpus-based tasks as significantly lower than the sentence-based tasks: $t(34) = -2.72$, $p = .010$ (two-tailed). No other significant contrasts were found. There may be some doubts whether students' respons-

es were not based on general preferences, likes and dislikes, rather than their actual judgment on effectiveness. Even if this were the case, however, their feedback may be relevant to considerations on the usefulness of various class activities. It is not very productive, or sensible, to implement teaching procedures to which students show much resistance or hostility. Fortunately, it appears from the students' responses discussed above that despite some differences, on the whole they had a positive attitude to what they did in their classes, and expected in effect to make progress. The differences, even if significant, do not indicate a negative assessment, but merely a less positive one.

4.11.6. Item 5

The next item in the survey enquired of students about their attitude to corpus-based activities in a straightforward way. They were given four Likert-scale options, from the least favorable opinion ("I did not like using such materials from the start and have not changed my mind" – briefly summarized in the table as "First not OK, no change"), to the most favorable one ("I liked using such materials from the start and have not changed my mind"). The in-between answers accounted for situations when the initial opinion (positive or negative) changed with time, as students had more opportunities to learn English by means of the DDL techniques. From the responses it can be gathered that generally the first impression was not favorable. Most of those initially unconvinced, however, later changed their minds, so they must have found corpus-based learning beneficial in some ways. A negative first impression is also reflected in the outcome of the experiment, as the lessons on *possibility* and *however* – the first ones in which these students had any experience of formal corpus-based teaching – had the poorest results. Before that, they did have an IT class where they found out what corpora are and how they could use them. This, it turns out, however, was not enough to prepare students for the use of DDL's innovative techniques in their grammar class. After that first lesson, however, the students' results for the corpus-based classes did improve, so the experience must have been positive on the whole.

Table 41. Responses to item 5

		Group		Total
		group 1+2	group 3	
Attitude	First not OK, no change	3	2	5
	First OK, then changed	2	0	2
	First not OK, then changed	10	6	16
	First OK, no change	3	8	11
Total		18	16	34

Again, the same proportion of students (ca. 20%) expressed their negative attitude to corpus-based teaching, the others being in favor of it either from the beginning or after some more experience of it. Reassuring as this can be, one needs to be aware of those students who found DDL a disappointment and balance corpus-based lessons with ones that this group would find more satisfying. One of the best students among the respondents, instead of choosing one of the options offered in answer to item 5, wrote her own response: "I know corpus-based activities are effective, but they are simply not my thing". Fair enough. The teacher needs to acknowledge that there are such students and accommodate their needs, at least to some extent. This can be done by involving them in a different type of task or adapting corpus based activities in such a way that they become more accessible to such students (some interactive or creative elements could be included, for example). Kaszubski (2011) labeled such students as "refusers" and also admitted that a small proportion of students could be expected to assume such an attitude. Those who accepted DDL activities willingly and performed their own independent corpus inquiries were categorized as "adopters", and the third group, the ones who did not reject the idea but did not find it particularly appealing, were called "minimalists".

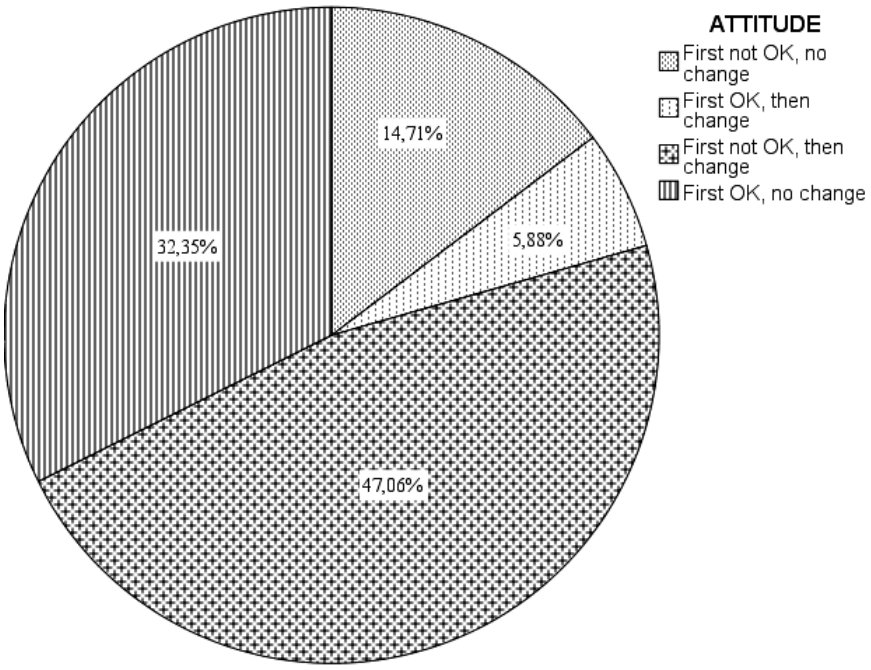


Figure 42. Pie chart for question 5 – attitude

4.11.7. Item 6

In item 6 students were asked to justify the opinion they expressed in the previous item by listing advantages and/or disadvantages of using corpus-based activities in the language classroom. Only five students decided not to list any advantages, and seven – disadvantages. This means that even the majority who expressed their approval in item 5 were aware of the shortcomings of these activities, which, however, must have been in their view outweighed by the benefits they offer. Although the item was open-ended, students' responses have been arranged into eight recurring categories on each side, so that some quantification would be possible.

DISADVANTAGES

The three most commonly listed problems with corpus-based teaching were that it was tiresome or unattractive, that the format was confusing and that it was time-consuming. Since students were free to mention as many features as they could think of,

the percentage values do not add up to 100% but indicate the proportion of students who listed particular features.

Table 42. Responses to item 6 – disadvantages of DDL activities

		Group		Total	Percent of students
		group 1+2	group 3		
Disadvantages	unattractive/monotonous	6	5	11	37.9%
	confusing format	5	3	8	27.6%
	time consuming	4	4	8	27.6%
	insufficient context	3	3	6	20.7%
	too much information	3	2	5	17.2%
	overwhelming	3	0	3	10.3%
	not creative	2	1	3	10.3%
	difficult	2	0	2	6.9%
Total		15	14	29	

The first disadvantage, monotonousness, is self-explanatory. It is not surprising that many students prefer language production activities to abstract analyses. This links with the less commonly mentioned, but also credible observation that DDL activities lack a creative aspect. They do involve students in building mental structures (rules), but are not creative in the more common understanding of the term – they do not give students opportunities to create new meanings and messages, and hence are less engaging. This is probably the most serious drawback. Also, many students find the very format of a concordance confusing: they complain that there is too much information cramped into too small a space, which has an off-putting effect due to which, students complain, they cannot focus on the task. Finally, some respondents said the concordances did not to offer a sufficient amount of context, as a result of which students had problems understanding how the form in question was actually used. This could be resolved by including a broader context, but then the previously mentioned charges would have been even more justified: more information would have been offered to the students to process, which would have made the task even more overwhelming – something of a vicious circle.

ADVANTAGES

Among advantages, the most commonly mentioned ones were the useful format and the real-life examples. The former, students reported, allowed them to observe regularities in language use more easily, which, incidentally, is exactly opposite to what was

listed among the top *dis*advantages above (*confusing format*). The latter is a quality that is often emphasized in advanced language courses: students want to (and should) learn the language as it is really used in written and spoken communication rather than its inauthentic, fabricated version. One fourth of the students listed DDL's *multitude of examples* as an advantage. Having studied theories of language acquisition, KJO students appreciate an enhanced amount of input that is relevant to their needs, and corpus-based materials offer just this. The last on the list is a characteristic that is close to what stands behind the concept of "real examples" but has been listed as a separate item here because there seems to be a difference between emphasizing that examples are *real*, and the fact that forms under scrutiny appear in their *natural context*. This relates to the information provided by BYU BNC, for example, about the type of text from which a concordance item originates (spoken, written, academic, etc.) as well as the tone and level of formality that can be recognized from such a small sample of language.

Table 43. Responses to item 6 – advantages of DDL activities

		Group		Total	Percent of students
		group 1+2	group 3		
Advantages	useful format	3	9	12	38.7%
	real examples	7	4	11	35.5%
	variety of examples	5	3	8	25.8%
	helps me improve	3	4	7	22.6%
	helps me understand	2	5	7	22.6%
	stimulates thinking (deduction)	3	4	7	22.6%
	helps me remember	2	4	6	19.4%
	natural context	4	1	5	16.1%
Total		15	16	31	

Apart from the characteristics of corpus-based materials listed above, students named various other beneficial effects DDL activities have on them: they help them improve their English, understand and remember the elements of language that are in question better, and, finally, they stimulate learners to analyze regularities and to form their own rules, which, some say, "works better" for them.

When both sets of responses are analyzed, it becomes clear that the same characteristics may be seen by some students as strengths and by others as weaknesses. This is the case with concordance layout, for example, and the assessment of the usefulness of the context provided in the form of a concordance. A lot depends on individual stu-

dents' preferences, and their cognitive and learning styles – the issues discussed in Chapter Three. The same can be said of other, more conventional classroom activities too. This is why it is important to use a variety of techniques in a language course, in balanced proportions, so that students' different needs could be satisfied at least to some extent.

Item 6 generated a variety of opinions and showed the students to be reflective learners, well aware of their needs and able to express their thoughts on what takes place in the classroom. Even if they had a generally favorable opinion of corpus techniques, they were able to recognize its weaker points, and the reverse – those who were not in favor of DDL activities, were still able to appreciate its advantages.

4.11.8. Item 7

Since the participants of the study were in a teacher-training program, they were also asked about their willingness to use DDL techniques in their future teaching, provided that they do become teachers. The responses were mostly affirmative: over 50% of students decided that they would include corpus-based activities in their prospective work as teachers. Over 30% decided that those activities could only be used with students who meet certain requirements: are at an advanced level, are in an ESP course or take one-on-one lessons. Sometimes the answers defined other qualities that such students would need to have: adult, highly motivated, or interested in analyzing grammar.

Table 44. Responses to item 7 – use of DDL techniques in future teaching

		Group		Total	Valid percent
		group 1+2	group 3		
Use in teaching in future	No	5	1	6	17.2%
	Only in advanced or other specialized classes (ESP, EAP, one-to-one, etc.)	6	5	11	31.4%
	Yes	7	11	18	51.4%
Total		18	17	35	100%

These responses confirm again students' positive experience with the corpus-based classes: they felt they were ready to implement similar activities in their own classes if /when they started teaching. It is commonly said that despite all the methodol-

ogy training, teachers often teach the way they were taught, so perhaps the above result should not be surprising; responses to the next item, however, undermine the validity of these declarations.

4.11.9. Item 8

The last but one item referred to students' own use of corpora when trying to resolve their doubts concerning English grammar and vocabulary. The responses cast some doubt on whether the intentions which the respondents declared concerning their use of corpora in teaching English have any chance of actually being fulfilled.

Table 45. Responses to item 8 – students' own use of corpora

		Group		Total	Valid percent
		group 1+2	group 3		
Own use of corpus	Never	9	4	13	37.2%
	Tried but was discouraged	0	2	2	5.7%
	Yes, only as homework	2	3	5	14.3%
	Yes, occasionally	3	8	11	31.4%
	Yes, regularly	4	0	4	11.4%
Total		18	17	35	100%

It is not probable that any of the 13 students who do not use a corpus at all would be willing, or able, to do so when preparing their lessons in the future. The more likely outcome seems to be that those four students who do refer to a corpus regularly, and some of the eleven who use it occasionally, could implement DDL techniques as teachers. The others may consider this as a possibility, but may not be motivated strongly enough to actually do it, if they do not use a corpus on a regular basis while they are still students.

The second possible response to the item was "I tried but was discouraged by something – (say what it was)". This was chosen by only two (2) respondents, but their answers reveal what the problem for many others could have been. One said that there were too many things to define in a corpus query, and the other that it is frustrating to her – probably for the same reasons. The online corpus that was presented to the students was BYU BNC, whose interface and operation are probably the easiest and most intuitive of all corpora available on the Internet, considering the very rich search options

it offers. Still, some settings need to be made for a search to give results relevant to the user's needs. Needless to say, formulating a search becomes easier with time, but few students are ready to spend such time learning how to use the tool.

4.11.10. Item 9

In the final item the students were asked for comments on their experience with corpus-based activities. Ten (10) of them decided to add their commentary, usually those who were most eager about the idea. Only one (1) person (a good student), chose to express her frustration with corpora, saying that they were "chaotic". Four (4) students declared that since they had started using corpora when writing essays, their grades had improved. One of them also added that she wished she had known about corpora earlier. Another student felt the need to excuse herself for her negative attitude and added that she wished to explore the subject further, pointing to authenticity of the materials as the major advantage. One respondent observed that the most difficult for her was the first corpus-based lesson (HOWEVER), but she could not tell whether the problem lay in the complexity of the subject or the fact that it was the first lesson of this kind, the novelty adding to the difficulty. The other responses emphasized the need to address corpus-based lessons to advanced, tertiary level students who are very language-oriented, and to use such materials and activities with moderation, as excess may be discouraging and tiring to learners.

4.12. Correlational study

The final part of this chapter is going to present and discuss some correlations that have been found among the subjects' various characteristics and opinions revealed in the survey. Their test results will also be set against some of these characteristics.

4.12.1. Mean gain values

For the sake of this analysis, some more variables have been calculated for the students participating in the study. The first two are supposed to represent individual students' overall progress resulting from the corpus lessons and from the conventional ones. The former has been labeled as the "corpus lessons mean gain" and the latter – the "conventional lessons mean gain." They have been arrived at by calculating the mean values of the gains from the three experimental lessons and three control lessons respectively. The correlations were calculated for each group separately, so as to avoid putting together values based on different sets of test questions. Even assuming that the lessons compared were of similar difficulty (as they were carefully planned to be), the mean gain values for the corresponding post-tests could not have been compared between groups. It also must be added that this statistic has been calculated only for those students who attended the experimental and control lessons at least twice, so that the measurement should not be based on one lesson only.

The third new variable was a gain value generated from the data obtained for one more lesson, performed with a DDL design for *all* students in the experiment. The class was devoted to the Saxon genitive and its various difficulties for Polish learners, connected with word order, noun modification, and problems with marking grammatical number (singular/plural) in the noun phrase. For details of the lesson, see page 446. The aim of this element of the study was to find one measure of performance resulting from corpus-based teaching that would embrace all participants at a time, and to find if those values correlated with any measurement of the student characteristics defined previously: exam results, attitude to corpus-based activities, language aptitude, etc. This indeed was the case: the gain values for the Saxon genitive (again calculated on the basis of 5 items in a pre-test/post-test design) were found to correlate with one variable in the data: students' grammar test results ($\rho = .37$, $p = .025$). The strength of the correlation is not very high, which is perhaps why it was not observed in the smaller samples (the two groups analyzed separately for their different corpus-based classes). The Spearman's rho non-parametric test of correlation had to be applied because of the non-normal distribution of gain values in the data for the Saxon genitive lesson. Admittedly, the reliability of a measurement based on one lesson is not very high, but at least there is an indication of a possible relationship between learners' accuracy (the construct measured by the

grammar tests) and the effectiveness of DDL techniques. The correlation indicates that students with a general higher accuracy could benefit from such lessons more than other students.

Such an outcome was initially expected of the correlational analysis; the assumption was that it would reveal a connection between students' test results for corpus-based classes and their proficiency in English (as represented by practical English test results and/or a mean result of the three major grammar tests administered during the course). The assumption was that DDL techniques may pose a challenge to weaker students because of the more difficult, authentic language that appears in corpus-derived materials, or because of the higher level of abstract, synthetic thinking that such techniques require. This did not occurr for the six lessons which constituted the main body of the experiment. Nor is there a significant correlation between the conventional lessons mean gains and students' test results. What is more, no correspondence has been found between the students' approval of corpus-based techniques and their mean gain results obtained in post-tests which followed lessons that used them. Such a lack of correlation does not render the lessons ineffective, but should prevent associating students' like or dislike of the techniques investigated with their success or failure in mastering the material introduced through them.

The lesson on the Saxon genitive, however, may be an indication that there is a connection between a learner's L2 proficiency, or to be precise, accuracy, and his or her success in corpus-based classroom activities, after all.

4.12.2. Students' attitudes to corpus-based techniques

Another set of correlations that was analyzed was correspondences between students' reported opinions of the three sets of teaching techniques (corpus-based, sentence-based, and text-based) and their other attributes identified in the study. First, it transpired – again contrary to expectations – that students' inclination for corpus-based learning correlates with that for sentence-based techniques (r = .42, p = .013, N = 35), but not with text-based techniques. One would think that in the case of both texts and concordances the key notion is authentic context, so it would be natural for students to see the two as more closely related than concordances and separate sentences. Still, it

seems that texts offer a less "economical" type of context, which requires a different kind of processing than either concordances or example sentences provided. It is not that text-based lessons are less popular, but it seems that generally there is little overlap between a preference for text-based teaching and a preference for DDL and sentence-based techniques.

Interestingly enough, there is a correlation between the preference for text-based teaching and the self-assessment of students' own level of proficiency in English: $\rho = .38$, $p = .023$, $N = 35$. This means that students who *declare* higher levels of proficiency tended to prefer text-based lessons. It may be that those students are more strongly motivated than others, and in their perception text activities, which require more extensive language processing, must be more beneficial than others. The principle of "no pain – no gain" may be involved here. This tendency is not, however, confirmed by any direct correlation with the exam or grammar test results. The only connection with any measure of the students' actual proficiency in English is indirect: a medium-size correlation between their self-assessment (the same factor as in the correlation discussed above) and practical English exam results ($\rho = .39$, $p = .027$, $N = 32$). This fact is in itself worth noting, as it indicates that the respondents were relatively accurate in defining their own proficiency in English. In this way a vague connection can be established between a preference for text-based activities and student's L2 proficiency, but more unequivocal evidence would be needed to confirm such a relation.

4.12.3. Task assessment

As stated above, in responses to item 4 students expressed their assessment of the effectiveness of various classroom activities. It is hardly surprising that values representing their evaluation of corpus-based tasks' effectiveness correlate with one another. The same is true of text-based activities. Table 46 presents all the correlations of this kind that have been found in the data:

Table 46. Correlations for same context-type tasks

Type of context	Correlated tasks	Spearman's rho (ρ)	Significance *p*	*N*
Corpus-based	Ns concordance vs. Parallel corpus search results	0.39	0.031	31
	NS concordance vs. Partial translation (conc)	0.44	0.015	30
	Parallel corpus search results vs. EC concordance	0.53	0.003	29
	Partial translation (conc) vs. Frequency lists	0.52	0.004	29
Text-based	Translation – fragments vs. Error correction - text	0.64	< 0.001	33

The correspondences in the effectiveness judgments on various corpus-based techniques are quite consistent: out of six different techniques listed in the item, five have been found to be part of at least one thus correlated pair. It may be said, therefore, that preferences for those techniques tend to occur in the same students. Similarly, assessment values for two out of the three text-based tasks correlate. The same, however, cannot be said about the sentence-based activities. It seems students did not perceive them as a group of tasks that had a common denominator: sentence-level processing; at least, this did not constitute a valid element of their evaluation.

Another observation can therefore be made on the basis of correlational analysis: students' judgments clustered around different types of activity in which a given task involved them rather than the form of presentation and the type and amount of context provided. Thus, correlations occur for various tasks involving translation, for example, regardless of whether it is performed in separate sentences, concordances or text fragments. The same is true for error correction tasks and writing. Table 47 presents all such correlations found in the data:

Table 47. Correlations for same activity-type tasks

Activity	Correlated tasks	Spearman's rho	Significance *p*	*N*
Translation	Partial transl. (concordances) vs. Sentences	0.40	0.024	32
	Sentences vs. Fragments	0.35	0.040	34
Error correction	Sentences vs. Text	0.43	0.015	32
	Sentences vs. Learner corpus concordances	0.36	0.052	29
Writing	Sentence building vs. Text building	0.64	< 0.001	33

Some other correlations have been found for various classroom activities in students' responses to the item on their effectiveness, but since the tasks involved do not share any characteristic features, it is rather difficult to account for these. For example, it is difficult to find an explanation for a relatively high correlation between students' evaluation of the effectiveness of a gapped concordance and of a sentence synthesis exercise ($\rho = .49$, $p = .005$, $N = 31$) or a gapped concordance and sentence-based error correction ($\rho = .42$, $p = .025$, $N = 29$). The only observation that can be made here is that again corpus-based tasks are perceived as more closely related to sentence-level activities than text-based ones. This division can be confirmed by the last correlation to be mentioned here, a negative one: that existing between building sentences with target forms and correcting errors in a text ($\rho = -.37$, $p = .034$, $N = 34$). The fact that it is a negative one means that the higher the evaluation of sentence building, the lower it was for text-based error correction. Again, the text processing has been set against sentence processing, which, on the other hand, is more strongly associated with corpus data processing.

4.12.4. General characteristics – correlations

Finally, a few interesting facts emerged from the correlational analysis as far as the participants' characteristics, preferences and test results are concerned. These are not directly connected with the central issues of this dissertation, but still may be of interest to ELT specialists.

It is often said, for example, that female language learners are more successful than male ones, and some research seems to confirm this (e.g. Peacock 1999). This is also reflected in the data collected for this study: there is a relatively strong correlation between students' grammar test mean results and female gender: $\rho = .41$, $p = .014$, $N = 36$. This has been confirmed in a t-test, which showed that the female students' mean result ($M_f = 65.1\%$, $SD_f = 0.91$) is significantly higher than the male students' ($M_m = 55.3$, $SD_m = 0.98$): $t(34) = 2.76$, $p = .009$, $\eta^2 = .18$. The size of the sample is far too small to make any generalizations, but still an outcome like this is worth noting. There was another interesting finding related to gender: as mentioned in section 4.12.2. above, the students' self-assessment correlated with their practical English examination

results, but when the same correlation was calculated for each gender separately, it turned out that it was sustained for the female students only ($\rho = .43$, $p = .029$, $N = 26$) and not for the male ones. It transpires that female students in this group had a much more accurate sense of their own advancement in learning English than their male colleagues, or at least were more open about defining it.

Another factor with which the end-of-year practical English examination results correlated was the grammar test means (calculated from three extensive tests administered during the grammar course): $r = .71$, $p < .001$, $N = 33$. Such a high correlation is hardly surprising, since grammar is evaluated in most components of the practical English examination, but the correlation itself, especially its high value, validates the two measurement instruments in a way, and makes them more dependable.

Finally, an interesting connection has been established between the exam results mentioned above and students' studies at another university department or at a different institution. Those who declared that they had completed or were still engaged in another course of studies tended to have higher practical English examination results than those who studied only at KJO ($\rho = .36$, $p = .045$, $N = 32$). Although not very strong, the correlation is quite interesting, considering that students are usually discouraged from enrolling in more than one course of studies, because of the organizational problems and excessive workload that entails. The above analysis of questionnaire responses as combined with test results indicates that this is not usually the case: such students are more strongly motivated, have more clearly defined aims, and are much better organized. Also, they may simply have a higher potential than those who have not undertaken any other studies. It may be worth taking into consideration, especially at recruitment, that such students can actually be more successful than those who decide to study only one subject.

Conclusions

The research project undertaken for this dissertation has been quite extensive: a learner corpus was built for its needs, two different error analysis studies were performed, a quasi-experiment was carried out, and finally a survey was administered and analyzed. Each element grew on the one/ones that came before, so the whole project should be

seen as a set of complementary elements. The first, traditional, error analysis was performed so as to provide clues for the corpus-based study. The learner corpus was used for the second error analysis, and then as one of the sources of teaching materials for the experimental lessons. These, in turn, were a necessary element for gathering the students' opinions on corpus-based teaching in the survey, whose results were analyzed against data obtained from the experimental study. In this way all the elements of the project became connected and contributed to one another.

The error analyses, first traditional and then computer-assisted, evidenced a wide variety of problems characteristic of advanced Polish learners of English. L1 interference is recognizable in all aspects of the written language in the data, from the most basic one (e.g. grapho-phonemic errors) to the more complex ones (e.g. clause combining errors). Every error in the essay database has been given a plausible explanation, which attributed its occurrence to a feature in students' L1. Out of seven categories formed for the purpose, three were especially frequently employed: *under-differentiation*, *calque*, and *feature absent from L1*. This could indicate the major sources of potential L1 interference problems in advanced learners, which should perhaps be considered by teachers in their day-to-day work with them. Early classroom intervention or focusing on these sources of confusion on a regular basis could prevent fossilization and help learners achieve better accuracy.

The corpus-based analysis has offered some insight into the frequency of accurate uses of a given form as compared with the inaccurate ones. What was found to be important was not the frequency of an item, but the proportion between its accurate and inaccurate uses in the corpus. At the advanced level students are expected to be able accurately to employ lower frequency words and phrases, which can help express themselves more precisely, and so absolute frequency was not the criterion of highest priority. If, however, the ratio between the total number of occurrences of an item and the number of errors among them was high, then that indicated that the language item needed more attention in the classroom. Using a corpus as a tool of error analysis proved particularly useful at the level of lexicogrammar. Predictably, searches for word choice problems gave especially interesting results, as corpus inquiry is best suited for such lexical issues. Hence, most of the experimental lessons addressed errors in this group (*possibility*, *state*, *approve*, and *amount*).

In terms of quantity, the analysis seemed to show that there are not many persistent errors that are shared by a clear majority of learners. Instead, there is a wide variety of different forms that divert from standard English, dispersed throughout the learner population. Targeting errors in class activities is therefore not the most efficient way of improving accuracy, whether through corpus activities or otherwise. The best use of corpora would be through students' individual work with them, in addressing their own doubts concerning English use and usage. The best effect of lessons like those performed in the study might be that they show learners what information can be obtained from corpora, and how these rich resources can be consulted, according to learners' own needs and interests. It is, therefore, a disappointment that, according to the survey, the percentage of students who declared to do so regularly is very small (11.4%), though quite a few stated that they refer to corpora occasionally (31.4%).

Corpus analysis has revealed some errors that can be found in native speakers' informal language as well. A good example here is the temporal use of what should be, but often is not, the Saxon genitive, e.g. **in five years time*. The form, though non-standard, appears very frequently in native written English, especially as informal as blog postings. It is important for students to realize that even though such forms are used by native speakers, they are not well received in formal, professional communication. Verifying L2 use through a corpus must always be done with an awareness of style and register, and not just through a straightforward frequency count. This is why corpora which provide information about source texts should be used, if available. Students need to understand that native speakers use non-standard language, too, and while it may be appropriate in some circumstances, in others it is not.

One needs to be aware of the technical limitations of corpus-based error analysis: it is best at scrutinizing how a given form is used by learners, but less effective in recognizing whether a target form has been chosen in an obligatory context. Defining such a context is often impossible in a corpus query, as it would require introducing abstract concepts rather than lexical choices or sequences of grammatically defined elements. For example, article omission in learner English can be identified with particular nouns of interest to the researcher or teacher, but it is impossible to define one search term to identify all article omission errors in the data, unless the corpus has a very deep annotation system that has a POS tag dedicated to countability in the noun. There is no such tag in the C7 tagset, currently in use with the official version of the BNC, for ex-

ample, created at the University Centre for Computer Corpus Research on Language of Lancaster University (UCREL). The limitations listed here should by no means discourage teachers and researchers from using learner corpora for error analysis and other diagnostic purposes. They offer valuable insight into the type of problems that a particular group of students copes with, and helps plan language courses in such a way that meets students' needs best, especially in terms of lexicogrammatical content. With the present-day accessibility of electronic text editing and communication, building ad-hoc corpora for groups of course participants within various action-research projects is attainable by every teacher. Such ad-hoc language corpora can be supplemented with more standardized and more carefully designed learner corpora like ICLE, for example, which have a more general appeal and higher credibility. Unfortunately, they are not easy to access for an average school teacher.

Further stages of the research project involved a quasi-experiment. The first stage proved inconclusive because of problems with the design of the study. Learner-related variables proved too difficult to control in the conditions of the first attempt and therefore the whole teaching component had to be repeated the following year, with most of the problems resolved. Already in this initial phase, when the first survey was carried out, some students' negative attitude to corpus-based activities became evident: they found such lessons confusing and demotivating, and this is thought to have affected their learning outcome. It became clear that the techniques analyzed were not a panacea, but had their enthusiasts and strong opponents. The consequence was that a particular group's results depended not only on what treatment they were offered, but also – to a large extent – on what were the individual preferences and capabilities of the students in them. The modified design of the study included a pre-test and post-test, and the questionnaires were no longer anonymous so that correlational studies could be performed. These were expected to reveal those variables which had a strong influence on the effectiveness of corpus-based teaching.

The pre-test carried out for the purposes of the experiment was based on translation technique, which was found to have significant advantages in studying interlanguage, especially with the focus on L1 interference: the activity itself depends heavily on L1 and therefore can provide focus on those areas of L1 which are particularly difficult for learners of English. Also, translation usually provides an obligatory context for a target form to be provided by the test-taker, and what is even more important,

the data obtained in this way is language production rather than recognition, which was crucial in the context of the error analysis. The drawbacks of the technique employed included some problems with assessment, as translation tasks are always partly open-ended and the researcher could not predict all the options that appeared in students' answers. Some of them, especially cases of avoidance, were simply difficult to fit into the rigid dichotomous scheme of the error analysis (*right* or *wrong*). Generally, however, the problems were limited by providing sentence frames, where possible, and the problem was reduced to a minimum.

The experimental (or quasi-experimental) element of the study has given rather unexpected results. The theoretical considerations regarding data-driven learning discussed in Chapters Two and Three allowed hypotheses that corpus-based learning activities are very effective and should give better results than conventional techniques, where exposure to language data is usually smaller, or at least less intensively focused on a given problem area. The data collected for this study, however, do not prove corpus-based techniques to be more effective. What is more, they emerge as less universal and more easily affected by features of the educational process which seem difficult to predict and control. These may be such learner-related features as learning style or characteristics of the subject matter (whether more lexical or more structural, for example). Especially disappointing was the first corpus lessons in each group – here the results were significantly better in the control group. It seems corpus-based learning requires some practice and familiarity before it gives positive results. This was certainly the case in this experiment. The later lessons did not show much difference in effectiveness; especially the ones devoted to problems of countability appeared to be very close in the outcome.

For most of the lessons in the project, delayed tests were performed apart from post-tests, so as to diagnose the durability of whatever effect a class technique had had. No consistent effect has been observed in this respect. An interesting thing that emerged from the delayed test, however, was the effect of the post-test results, especially if they were very low. Then, regardless of the technique employed in the lesson, the outcome of the delayed test was significantly higher in some cases. This shows the importance of testing not only as a diagnostic tool for the teacher, but also as a motivating instrument for the student. Interestingly, when the test results were not bad, the delayed tests showed slightly lower scores, indicating perhaps that students did not have any reason

to revise the problem and were beginning to forget what they had learned. Generally the four delayed tests did not reveal any consistent differences in how durable the effects of teaching were either through a corpus or through ordinary forms of instruction.

The analysis of the results for the two sets of three lessons of the same type did not reveal any significant differences either. What did transpire, though, was the differences between the two groups, again. One of them turned out to be simply more "teachable" than the other, no matter what technique was employed. The small differences in effectiveness that were found were minor and failed statistical verification. The groups that were instructed in the more traditional ways had more normally distributed results, and their gain values were slightly higher. The result for one of the two sets of lessons was very close to the margin of statistical significance, to the benefit of the conventional teaching.

The final point in the discussion of the experiment must refer to the difficulty of controlling the relevant variables in a typical educational setting, starting from the selection of subjects and dividing them into groups, which is normally done outside a researcher's control, and ending with problems like absenteeism and student attrition. The number of students available for the experiment was also limited by the situation of the educational institution the researcher had access to. On the whole, the consistency of this study has been affected by all these factors, and so its findings should be verified under optimal experimental conditions, with the random selection of subjects into groups and a bigger number of subjects to examine.

Unlike the experimental part, the survey has given a much more positive feedback on corpus-based teaching and learning. All the questions about students' opinions on corpus lessons received very positive responses from a big majority of the students. The most commonly chosen adjective that described corpus-based techniques was *thought-provoking* (30 students out of 35 chose it). Interestingly enough, among the positive adjectives, the least commonly selected one was *effective* (23 out of 35), which is relevant to the research question of this study. From responses to this and other items it may be concluded that most students find corpus-based activities interesting and useful, but are not very confident about their effectiveness. Among the negative adjectives, the most common, though from very few choices were *overwhelming* and *time-consuming* (7 out of 35 each). Considering the relatively small number of these answers, however, DDL techniques can be said to have been well received by the respondents.

This is confirmed by responses to another question, the one that asks students directly how they felt about corpus-based activities. The response was generally positive, although it must be noted that most students were not very enthusiastic at first, and admitted that initially they did not like these classes very much.

When asked about the disadvantages of DDL, students most often referred to them as being monotonous and generally unattractive. Quite a few complained about the concordance format, saying that it provided too much information, and therefore confused them. Others considered these tasks to be too time-consuming. Ironically, the 'useful' concordance format was listed by students most often as an *advantage* of corpus materials, too. Other commonly mentioned advantages were the authenticity of the examples and their variety. It seems, therefore, that there is a whole spectrum of opinion here, and the decisive factor here is the individual preference of the learner. The fact that there were around 20% of respondents who declared a strong dislike of DDL techniques indicates that such techniques should be used in moderation; also, they could be either supplemented or adapted in such a way as to minimize those students' frustration in class.

In the item referring to the effectiveness of three groups of activities: corpus based, sentence-based, and text-based, the opinions about the first one were not evenly distributed between the two groups: there were generally fewer corpus enthusiasts in group 1+2 than in group 3, and a few of the students in the former declared very critical opinions of corpus activities, so much so that the difference of opinion between the groups was significant. Such differences were not observed for the other types of activities listed in the item (sentence-based or text-based), so it seems that corpus-based lessons were to some extent controversial. All the same, the criticism expressed by a few students must not be seen as an overall negative evaluation. What needs to be said is that students assessed the effectiveness of DDL significantly less favorably than that of sentence-based classroom activities. As for text based techniques, their effectiveness was not seen as significantly better or worse. It seems that despite all the benefits that context provides, the respondents found isolated sentences clearer and more useful, after all.

Despite the generally positive attitude to corpus-based activities in class, few students actually use corpora on their own, and if they do, it is only occasionally. Only four respondents declared referring to a corpus regularly, and it must be said they were

very enthusiastic about it. They reported having made progress in their writing, confirmed by better grades in their writing course. Most students, however, were discouraged by the many options one is required to define for a good corpus query, and found working with corpora frustrating. It seems they were not motivated strongly enough to persist. With classroom use of corpora having proved in some ways controversial, individual use should be encouraged and promoted, as it is through addressing their individual L2 problems that students can benefit from corpus reference most. Interestingly, half of the respondents declared intentions of planning to use corpora in their own teaching. This statistic must be then approached with prudence: if they are not using corpora to improve their own accuracy in English, it is not very likely that they will do so for their future students' sake.

On the whole the survey provided a positive feedback from the participants of the experiment, most of whom found the DDL lessons at least an interesting diversion from the normal course of things. A few found them an inspiration for exploring corpus resources on their own. Some interesting comments were offered about the advantages and disadvantages of corpus-based teaching, and about the experience of such classes. The heaviest charge against DDL activities was that they do not engage students in any creative process, thus making the lesson tiring and monotonous. These opinions were expressed by a minority, but still are worth considering, along with the generally positive, sometimes even enthusiastic attitude of most students. When interpreting the answers to the survey, the author had to bear in mind that the questionnaire was not anonymous. Both the introduction to the survey and oral instruction given immediately before the survey was distributed, insisted that the answers should be objective and honest, and that the students would not be doing the teacher any service by giving insincere responses to just make her feel good. The questionnaire was administered after all grades had been decided, so that there would be no fear of possible bias on the part of the teacher in reaction to critical responses. Still, the lack of anonymity must be considered as a possible source of some 'noise' in the data.

Finally, the results of the survey and the experimental study were combined in a correlational analysis. The expectation had been that some characteristics of the students identified in the survey – level of proficiency, attitude to DDL or language aptitude, for example – would correlate with the gain values obtained in the test data analysis. Only one such correlation was found, for an additional set of data that embraced all

the students with one variable, namely, the gain values for a corpus-based lesson on the Saxon genitive, a form in English grammar that is particularly challenging for Polish learners. These gain values were found to be in significant correlation, though moderate in its power effect, with students' grammar test results. From this one could conclude that those learners who do better in grammar and have a higher degree of accuracy in L2 gain greater benefit from classes like the one on the Saxon genitive than students who are less successful in grammar. This observation, however, would have to be confirmed by more extensive studies, with more data to analyze.

Conclusion

The major aim of this thesis was to put to test corpus-based language-teaching techniques, especially in the context of remedial instruction. There were some grounds to suppose that the innovative classroom procedures and materials connected with corpus-based language teaching techniques, with their powerful visual impact and high credibility based on statistical data, would outperform the more conventional language teaching techniques in terms of their effectiveness. These expectations proved to be overly optimistic. It is not that the DDL techniques were found ineffective, by any means, but they are clearly no panacea either. They proved most successful in lessons on making good lexical choices and using words in different contexts, while teaching grammar-related issues turned out to be less beneficial. Therefore, the blending of grammar and lexis into the broad category of lexicogrammar, which was postulated for the needs of error classification, may not always be advisable. Apparently, some distinction between the two may still be useful when considering what type of instruction would be best for a given language problem.

The corpus tools involved in the study were not limited to providing input to students. A key element of the design was building a learner corpus and using it to diagnose learners' problems and areas of difficulty. This part of the project could serve as a proof for other teachers that this can be done, at no cost and to the great benefit of both the teacher, who receives valuable data on his/her students, and the students themselves. They receive feedback on their problems and are given an opportunity to learn from their own mistakes. This part of the study was also the most satisfying to the author, and will definitely be further developed and expanded to include other types of students – at different levels of proficiency or with different backgrounds, for example.

As far as DDL techniques are concerned, they did seem very convincing to some of the students, and they grew to become regular corpus users after they had been introduced to the idea. There were many learners who had positive attitudes towards DDL techniques, but it must be said that some found them rather daunting. What is surprising is that learners' attitudes towards corpus-based classroom activities did not seem to be significantly correlated with the effectiveness of such instruction for those students. It seems that other learner-dependent factors such as aptitude or motivation may have played a more decisive role here. As already stated, a lesson's effectiveness depends not only on its content and form but also on a variety of other learner-specific factors that are beyond an average teacher's recognition and control.

It must be said that effectiveness is not the only criterion to be depended upon when choosing a classroom procedure. As long as it is generally effective, a technique may be worth recommending for many other reasons. DDL offers numerous advantages: it increases learners' interest in the target language by making them ask more questions, it encourages students to take the initiative and become more autonomous learners, and, what is most important, it offers a powerful instrument to do this. When informed of the results of this study, Boulton (p.c.) commented as follows: "[T]he point is not to 'prove' that DDL is better than anything else and should be used exclusively, but that if it's useful for some things for some people some of the time, then it's useful as an extra tool in the teacher's arsenal". It appears that this is what this thesis has been able to achieve: to present the ways in which DDL can be *useful* when confronting advanced learners' interference errors, and the benefits it offers to language teachers and their students. The corpus-based activities introduced a different dynamic to the classroom, in that students became more aware of the flexibility of language, the variety of options available, and the criteria of choosing one form rather than another. They became less focused on the categorical distinctions between *correct* and *incorrect* choices, and more interested in the sometimes subtle differences in meaning that result from those choices.

What is more, the fact that the results of the experiment were rather inconclusive does not prove beyond doubt that DDL was not more effective in the end. Language learning is a very complex process, which does not always progress in a linear fashion. The processing of input offered to students in the experimental lessons may have required more time before the information provided could become employed by the learn-

ers in their active use of the target language. The restructuring of the patterns which learners had used beforehand may have taken longer than the experiment allowed. The remedial function of the experimental lessons may have had a strong impact: some of the errors may have been more fossilized than others, which could have affected the final outcome. Other factors may have played a role as well. Generally, many studies on the effectiveness of teaching techniques suffer from the same problem: the complexity of the teaching situation makes it very difficult to isolate the one variable (type of instruction) from all other factors which contribute to the outcome of the experiment. Recent developments in research methodology might offer a solution to this conundrum.

Larsen-Freeman and Cameron (2008) as well as Ellis and Larsen-Freeman (2009) advise introducing the fairly new methodology of complex adaptive systems into applied linguistics. It is believed to be able to account for the constant change that the object of examination, i.e. interlanguage, undergoes; change is actually a key concept to the method. The traditional approach isolates dependent variables which are believed to be part of the cause-and-effect chain in language development, and assumes the others to be independent, external to the processes under scrutiny, or fully controlled. Complexity theory does not assume the existance of dependent and independent variables, but proposes to choose a focal point for study, and treat other factors as background, which still undergo change and can affect the outcome of a study. According to Larsen-Freeman and Cameron (2008: 235) "we need to remain open to seeking explanations for observations from outside of the focal subsystem". The crucial decision is the choice of focus, and how broad it is. The common fault of the traditional, "reductionist" methodology is the "ecological invalidity" of its outcomes, due to the underappreciation of the context in which the study is conducted, both at the stage of the experiment and later, in the interpretation of its results.

There are several alternative modes of research recommended within the theory of complex systems. In *formative experiments*, the researcher defines a pedagogical goal and tries to find the ways in which it can be achieved. *Design experiments* involve introducing repeated changes to the learning environment, with the researcher observing their effects. Another suggestion is a "longitudinal, case-study, time-series apporach, which enables connections to be made across levels and timescales" (Larsen-Freeman and Cameron 2008: 245). Other suggested techniques include microdevelopment studies, computer modeling, brain imaging, and… corpus studies. Ellis and Larsen-Freeman

(2009) published a collection of papers documenting practical uses of the complex adaptive system methodology in applied linguistics. These recommendations are not always welcome by more traditionally-oriented linguists as these research methods are often seen as less 'scientific'. It must be said, however, that complexity theory is becoming more and more popular in such fields as economics or medicine, where reliability is, after all, highly valued. It may be that the effectiveness of foreign language teaching in general, and data driven learning techniques in particular, could be considered from these novel perspectives as well, and such studies could give more conclusive results.

The outcome of this project indicates that more research needs to be done into the issues of DDL effectiveness. The overall positive response from the majority of the students involved in it suggests that corpus-based language teaching techniques are generally welcome, though not without some reservations. They seem more effective in dealing with lexical rather than grammatical aspects of language, but this observation requires more analysis in further research. Drawing the line between the two when referring to corpora is still a challenge.

Even if the measurable effectiveness is not proven to be higher for DDL, its techniques are well worth promoting for classroom use because of their life-long learning effect. Learners develop a skill of using a corpus for reference, and become accustomed to this form of language data. Thus they are given an opportunity to become more confident and more independent learners, especially after they have completed their formal language education. With the increasing online availability of language corpora, learners have access to a resource that can give them confidence in their use of the target language, provided that they know how to use it. This is why corpus-based lessons should be included in language courses, even if their immediate effectiveness is not proven to be higher than that of other types of instruction.

References

Adolphs, Svenja. 2006. *Introducing electronic text analysis: A practical guide for language and literary studies*. London: Routledge.

Agnello, Francesca. 1977. "Exploring the pidginization hypothesis: a study of three fossilized negation systems", in: Carol Henning (ed.), *Proceedings of the Los Angeles Second Language Research Forum.* Los Angeles: English Department (TESL), University of California at Los Angeles, 246-261.

Allan, Rachel. 2006. "Data-driven learning and vocabulary: Investigating the use of concordances with advanced learners of English", *Centre for Language and Communication Studies, Occasional Paper* 66. Dublin: Trinity College Dublin.

Bachman, Lyle F. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, Lyle F. 2004. *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.

Bachman, Lyle F. 2005. *Statistical analyses for language assessment*. *Workbook and CD-ROM*. Cambridge: Cambridge University Press.

Baker, Paul. 2010. *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh University Press.

Barlow, Michael. 2000. "Parallel texts in language teaching", in: Simon Botley, Tony McEnery and Andrew Wilson (eds.), *Multilingual corpora in teaching and research*. Amsterdam: Rodopi, 106-115.

Barlow, Michael. 2002. Corpora, concordancing, and language teaching. Proceedings of the 2002 KAMALL International Conference, Daejon, Korea

(http://www.itfind.or.kr/Report01/200401/IITA/IITA-0766-014/IITA-0766-014.pdf) (date of access 29 Nov. 2011).

Barlow, Michael. 2011. "Corpus linguistics and theoretical linguistics", *International Journal of Corpus Linguistics* 16, 1: 3-44.

Barlow, Michael and Stephanie Burdine. 2007. *American phrasal verbs*. (CorpusLAB Series). Houston, TX: Athelstan.

Barnbrook, Goeff. 1996. *Language and computers. A practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.

Batstone, Rob. 1994. *Grammar*. Oxford: Oxford Univrsity Press.

Biber, Douglas. 1993. "Using register-diversified corpora for general language studies", *Computational Linguistics* 19, 2: 219-241.

Biber, Douglas, Susan Conrad and Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman grammar of spoken and written english*. Harlow: Pearson Education Limited.

Bley-Vroman, Robert. 1983. "The comparative fallacy in interlanguage studies: The case of systematicity", *Language Learning* 33: 1-17.

Bley-Vroman, Robert. 1989. "What is the logical problem of foreign language learning?", in: Susan M. Gass and Jacquelyn Schachter (eds.), *Linguistic perspectives on second language acquisition.* Cambridge: Cambridge University Press, 41-68.

Bolinger, Dwight. 1976. "Meaning and memory", *Forum Linguisticum* 1, 1: 1-14.

Bonk, Curtis Jay and Donald J. Cunningham. 1998. "Searching for learner-centered, constructivist and sociocultural components of collaborative educational learning tools", in: Curtis Jay Bonk and Kira S. King (eds.), *Electronic collaborators: Learner-centered technologies for literacy, apprenticeship, and discourse*. Mahwah, NJ: Lawrence Erlbaum Associates Inc. Publishers.

Boulton, A. 2007. "But where's the proof? The need for empirical evidence for data-driven learning", in: M. Edwardes (ed.), *Proceedings of the BAAL Annual Conference 2007*. London: Scitsiugnil Press, 13-16. (http://hal.archives-ouvertes.fr/docs/00/32/67/04/PDF/2007_boulton_BAAL_proof.pdf) (date of access: 31 May 2012).

Boulton, Alex. 2008a. "DDL: Reaching the parts other teaching can't reach?", in: Ana Frankenberg-Garcia (ed.), *Proceedings of the 8th Teaching and Language Corpora Conference*. Lisbon, Portugal: Associação de Estudos e de Investigação Cientifíca do ISLA-Lisboa, 38-44.

Boulton, Alex. 2008b. "Looking for empirical evidence of data-driven learning at lower levels", in: Barbara Lewandowska-Tomaszczyk (ed.), *Corpus linguistics, computer tools, and applications: State of the art*. Frankfurt: Peter Lang, 581-598.

Boulton, Alex. 2009a. "Testing the limits of data-driven learning: language proficiency and training", *ReCALL* 21, 1: 37-51.

Boulton, Alex. 2009b. "Tim Johns' data-driven learning (DDL)", (http://arche.univ-nancy2.fr/file.php/967/2009_Johns_summaries.pdf) (date of access: 20 June 2012).

Boulton, Alex. 2009c. "Corpora for all? Learning styles and data-driven learning", in: Michaela Mahlberg, Victorina González-Díaz and Catherine Smith (eds.). *Proceedings of the Corpus Linguistics Conference CL 2009. University of Liverpool, UK. 20-23 July 2009*. (http://ucrel.lancs.ac.uk/publications/cl2009/150_FullPaper.doc) (date of access: 27 Dec. 2011).

Boulton, Alex. 2010a. "Data-driven learning: Taking the computer out of the equation", *Language Learning* 60, 3: 534–572.

Boulton, Alex. 2010b. "Learning outcomes in corpus consultation", in: María Moreno Jaén, Fernando Serrano Valverde and María Calzada Pérez (eds.), *Exploring new paths in language pedagogy: Lexis and corpus-based language teaching*. London: Equinox, 129-144. (http://hal.archives-ouvertes.fr/index.php?halsid=qm4ucp8f1cjta9sb0g4cvkhc03&view_this_doc=hal-00502629&version =2 – author manuscript) (date of access: 7 feb. 2012).

Boulton, Alex. 2011a. "Data-driven learning: The perpetual enigma", in: Stanisław Goźdź-Roszkowski (ed.), *Explorations across languages and corpora*. Frankfurt: Peter Lang, 563-580.

Boulton, Alex. 2011b. "Empirical research in data-driven learning: A summary", (http://arche. univ-nancy2.fr/file.php/967/a_DDL_empirical_survey_2011_Sept.pdf) (date of access: 27 Dec. 2011)[32].

[32]This publication is a web supplement to Boulton (2010b)

Boulton, Alex. 2012. “Beyond concordancing: Multiple affordances of corpora in university language degrees”, *Procedia. Social and Behavioral Sciences* 34: 33-38.

Brown, James Dean. 1988. *Understanding research in second language learning: A teacher’s guide to statistics and research design*. Cambridge: Cambridge University Press.

Brown, Roger and Camille Hanlon. 1970. “Derivational complexity and order of acquisition in child speech”, in: Roger Brown (ed.), *Psycholinguistics*. New York: Free Press.

Burt, Marina K. and Carol Kiparsky. 1972. *The gooficon. A repair manual for English*. Rowley, MA: Newbury House.

Burt, Marina K. and Carol Kiparsky. 1974. “Global and local mistakes”, in: John H. Schumann and Nancy Stenson (eds.), *New frontiers in second language learning*. Rowley, MA: Newbury House, 71-80.

Bybee, Joan. 2006. “From usage to grammar: the mind’s response to repetition”, *Language* 82, 4: 711-733.

Canale, Michael and Merrill Swain. 1980. “Theoretical bases of communicative approaches to second language teaching and testing”, *Applied Linguistics* 1, 1-47.

Carter, Ronald and Michael McCarthy. 2006. *Cambridge grammar of English*. Cambridge: Cambridge University Press.

Celce-Murcia, Marianne and Jacquelyn Schachter. 1977. “Some reservations concerning error analysis”, *TESOL Quarterly* 11, 4: 441-451.

Chan, Tun-pei and Hsien-Chin Liou. 2005. “Effects of web-based concordancing instruction on EFL students’ learning of verb-noun collocations”, *Computer Assisted Language Learning* 18, 3: 231-251.

Chomsky, Noam. 1959. “Review of B.F. Skinner’s *Verbal behavior*”, *Language* 35, 1: 26-58.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1981. (IFA 7th edition, 1993) *Lectures on government and binding*. Berlin: Mouton de Gruyter.

Christiansen, Morten H. and Nick Chater. 2001. “Connectionist psycholinguistics in perspective”, in: Morten H. Christiansen and Nick Chater (eds.), *Connectionist psycholinguistics.* Westport, CT: Ablex Publishing, 19-75.

Chun, Ann, Richard Day, N. Ann Chenoweth and Stuart Luppescu. 1982 “Errors, interaction, and correction: A study of native-non-native conversations”, *TESOL Quarterly* 16, 4: 537-547.

Cobb, Tom. 1997. “Is there any measurable learning from hands-on concordancing?”, *System* 25, 3: 301-315.

Cobb, Tom. 1999a. “Applying constructivism: a test for the learner-as-scientist”, *Educational Technology Research and Development* 47, 3: 15-31. (http://www.lextu-tor.ca/cv/pdf/applying_constructivism.pdf) (date of access: 1 March 2012).

Cobb, Tom. 1999b. “Breadth and depth of lexical acquisition with hands-on concordancing”, *Computer Assisted Language Learning* 12, 4: 345-360.

Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences*. (2nd edition.) Hillsdale, NJ: Erlbaum.

Collins German English dictionary. 2011. (http://www.collinsdictionary.com/ dictionary/german-english) (date of access: 27 Feb. 2012).

Conrad, Susan M. 1999. “The importance of corpus-based research for language teachers”, *System* 27: 1-18.

Cook, Guy. 2010. *Translation in language teaching: An argument for reassessment*. Oxford: Oxford University Press.

Cook, Vivian. 1993. *Linguistics and second language acquisition*. London: Macmillan.

Corder, Stephen Pit. [1967] 1981. “The significance of learners’ errors”, *Error analysis and interlanguage.* Oxford: Oxford University Press, 5-13. [Originally published, Stephen Pit Corder, *International Review of Applied Linguistics*, 5: 160-170.].

Corder, Stephen Pit. [1971] 1981. “Idiosyncratic dialects and error analysis”, *Error analysis and interlanguage.* Oxford: Oxford University Press, 14-25. [Originally published, *International Review of Applied Linguistics*, 9: 149-159.].

Corder, Stephen Pit. [1976] 1981. “The study of interlanguage”, *Error analysis and interlanguage.* Oxford: Oxford University Press, 65-78. [Originally published, *Proceedings of the Fourth International Congress of Applied Linguistics.* Munich: HochschulVerlag. 122-127.].

Corder, Stephen Pit. [1983] 1992. “A role for the mother tongue”, in: Susan M. Gass and Larry Selinker (eds.), *Language transfer in language learning*. (revised edition.) Amsterdam: John Benjamins Publishing Company. 18-31.

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, teaching, assessment* [*CEFR*]. Cambridge: Cambridge University Press.

Coxhead, Averil. 2000. “A new academic word list”, *TESOL Quarterly* 34, 2: 213–238.

Crocker Matthew W. and Frank Keller. 2006. “Probabilistic grammars as models of gradience in language processing”, in: Gisbert Fanselow, Caroline Féry, Ralf Fogel and Matthias Schlesewsky (eds.), *Gradience in grammar. Generative perspectives*. Oxford: Oxford University Press, 227-245.

Crystal, David. 1995. *The Cambridge encyclopedia of the English language*. Cambridge: Cambridge University Press.

Dagneaux, Estelle, Sharon Denness and Sylviane Granger. 1998. “Computer-aided error analysis”, *System: An International Journal of Educational Technology and Applied Linguistics* 26: 163-174.

Davies, Eirlys. 1983. “Error evaluation: the importance of viewpoint”, *ELT Journal* 37, 4: 304-311.

Davis, Boyd and Lisa Russell-Pinson. 2002. “Concordancing and corpora for K-12 teachers: Project MORE”, in: Ulla Connor and Thomas A. Upton (eds.), *Applied corpus linguistics. A multidimensional perspective*. (Language and Computers 52.) Amsterdam: Rodopi. 147-169. (http://www.ingentaconnect.com/content/rodopi/lang/2004/00000052/00000001) (date of access 26 Nov. 2010)

De Cock, Sylvie and Magali Paquot. 2010. “The monolingual learners’ dictionary as a productive tool: the contribution of learner corpora”, in: Mari Carmen Campoy, Begona Belles-Fortuno and Maria Lluisa Gea-Valor (eds.), *Corpus-based approaches to English language teaching*. London: Continuum, 195-204.

Dörnyei, Zoltán. 2007. *Research methods in applied linguistics*. Oxford: Oxford University Press.

Droździał-Szelest, Krystyna. 1997. *Language learning strategies in the process of acquiring a foreign language*. Poznań: Motivex.

Dulay, Heidi and Marina Burt. 1974. “Errors and strategies in child second language acquisition”, *TESOL Quarterly* 8, 2: 129-136.

Dulay, Heidi, Marina Burt and Stephen Krashen. 1982. *Language two.* Rowley, MA: Newbury House.

Dušková, Libuše. 1969. "On sources of errors in foreign language learning", *International Review of Applied Linguistics* 7: 11-36.

Dziubalska-Kołaczyk, Katarzyna, Anna Bogacka, Dawid Pietrala, Mikołaj Wypych and Grzegorz Krynicki. 2006. *PELT: An English language tutorial system for Polish speakers* (http://ifa.amu.edu.pl/cslp/downloads/cslp_amu_d2006a01_multiling paper.pdf) (date of access: July 24, 2010).

Eckmann, Fred R. 1977. "Markedness and the contrastive analysis hypothesis", *Language Learning* 27, 2: 315-330.

Elliot, Andrew J. and Particia G. Devine. 1994. "On the motivational nature of cognitive dissonance as psychological discomfort", *Journal of Personality and Social Psychology* 67, 3: 382-394.

Ellis, Nick. 2002. "Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition", *Studies in Second Language Acquisition* 24, 2: 143-188.

Ellis, Rod. 2001. "Introduction: Investigating form-focused instruction", in: Rod Ellis (ed.), *Form-focused instruction and second language learning*. Malden, MA: Blackwell Publishers, 1-46.

Ellis, Rod. 2008. *The study of second language acquisition*. (2nd edition.) Oxford: Oxford University Press.

Ellis, Rod. 2010. "Explicit Form-focused instruction and second language acquisition", in: Bernard Spolsky and Francis M. Hult (eds.), *The handbook of educational linguistics*. Oxford: Wiley-Blackwell, 437-455.

Ellis, Rod, Helen Basturkmen and Shawn Loewen. 2001. "Learner uptake in communicative ESL lessons", *Language Learning* 51, 2: 281-318.

Ellis, Rod and Gary Barkhuizen. 2005. *Analysing learner language*. Oxford: Oxford. University Press.

Ellis, Rod, Shawn Loewen and Rosemary Erlam. 2006. "Implicit and explicit corrective feedback and the acquisition of L2 grammar", *Studies in Second Language Acquisition* 28, 2: 339-368.

Ervin-Tripp, Susan M. 1974. "Is second language learning like the first", *TESOL Quarterly* 8, 2: 111-127.

Faerch, Claus and Gabriele Kasper. 1983. “Plans and strategies in foreign language communication”, in: Claus Faerch and Gabriele Kasper (eds.), *Strategies in interlanguage communication*. London: Longman, 20-60.

Felder, Richard M. and Linda K. Silverman. 1988. “Learning and teaching styles in engineering education”, *Engineering Education*, 78, 7: 674-681. (New preface: 2002.) (http://www4.ncsu.edu/unity/lockers/users/f/felder/public/Papers/LS-1988.pdf) (date of access: 1 Sept. 2012).

Flowerdew, Lynne. 2008. “Pedagogic value of corpora: a critical evaluation”, in: Ana Frankenberg-Garcia (ed.), *Proceedings of the 8th Teaching and Language Corpora Conference*. Lisbon: Associação de Estudos e de Investigação Cientifíca do ISLA–Lisboa, 115-119.

Francis, W. Nelson and Henry Kucera. 1979. *Brown corpus manual*. (revised edition; 1st edition 1964.) Providence, RI: Brown University. (http://icame.uib.no/brown/bcm.html#bc1) (date of access: 3 Jan. 2012).

Friedl, Jeffrey E. F. 2001. *Wyrażenia regularne* [Regular expressions]. Gliwice: Helion, O’Reilly.

Fries, Charles C. 1945. *Teaching and learning English as a foreign language*. Ann Arbor, MI: University of Michigan Press.

Gabrielatos, Costas. 1999. “Inference: Procedures and implications for TEFL. Part 1: Background”, *TESOL Greece Newsletter* 63: 15-20 (http://www.gabrielatos.com/InferenceTEFL-1.pdf) (date of access: date of access: 5 Aug. 2012).

Gabrielatos, Costas. 2005. “Corpora and language teaching: Just a fling or wedding bells?”, *TESL-EJ* 8, 4: 1-35 (http://eprints.lancs.ac.uk/530/1/TESL-EJ_8(4)-A1.mht) (date of access: 5 Aug. 2012).

Gahl, Suzanne and Alan C. L. Yu. 2006. “Introduction to the special issue on exemplar-based models in linguistics”, *The Linguistic Review* 23: 213–216.

Gass, Susan. 1997. *Input, interaction, and the second language learner*. Mahwah, NJ: Lawrence Erlbaum.

Gass, Susan and Larry Selinker. 1994. *Second language acquisition: An introductory course*. Hillsdale: Lawrence Erlbaum Associates.

Gass, Susan and Larry Selinker. 2008. *Second language acquisition: An introductory course*. (3rd edition.) New York: Routledge.

Gerngross, Günter, Herbert Puchta and Scott Thornbury. 2007. *Teaching grammar creatively*. Cambridge: Cambridge University Press

Ghadessy, Mohsen. 1980. "Implications of error analysis for second/foreign language acquisition", *International Review of Applied Linguistics* 18, 2: 94-104.

Granger, Sylviane. 1997. "On identifying the syntactic and discourse features of participle clauses in academic English: native and non-native writers compared", in: Jan Aarts, Inge de Mönnink and Herman Wekker (eds.), *Studies in English language and teaching*. Amsterdam: Rodopi, 185-198.

Granger, Sylviane. 1998a. "The computer learner corpus: a versatile new source of data for SLA research", in: Granger, Sylviane (ed.), *Learner English on computer*. London: Longman.

Granger, Sylviane. 1998b. "Prefabricated patterns in advanced EFL writing: Collocations and lexical phrases", in: Anthony Paul Cowie (ed.), *Phraseology: Theory, analysis and applications.* Oxford: Clarendon Press, 145-160.

Granger, Sylviane. 2003. "The International Corpus of Learner English: A new resource for foreign language learning and teaching and second language acquisition research", *TESOL Quarterly* 37, 3: 538-546.

Granger, Sylviane and Hellen Swallow. 1988. "Flase friends: A calleidoscope of translation difficulties", *Langage et I'Homme* 23: 108-120. (http://dial.academielouvain.be/vital/access/services/Download/boreal:75703/PDF_01) (date of access 29 July 2012).

Granger, Sylviane, Estelle Dagneaux and and Fanny Meunier (eds.). 2002. *The international Corpus of Learner English: Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain.

Griffiths, Carol and Zhou Chunhong. 2008. "Researching error correction in China: Procedure, product and pitfalls", in: Mirosław Pawlak (ed.), *Investigating English language learning and teaching*. Kalisz: Uniwersytet im. Adama Mickiewicza, 127-138.

Hadley, Gregory. 2002. "Sensing the winds of change: an introduction to data-driven learning", *RELC Journal* 33, 2: 99-124.

Halliday, Michael Alexander Kirkwood. 1973. *Explorations in the function of language*. London: Edward Arnold.

Halliday, Michael Alexander Kirkwood. 1985. *Introduction to functional grammar*. London: Edward Arnold.

Hammerly, Hector. 1991. *Fluency and accuracy: Toward balance in language teaching and learning*. Clevedon: Multilingual Matters LTD.

Hatch, Evelyn and Hossein Farhady. 1982. *Research design and statistics for applied linguistics.* Rowley: Newbury House Publishers, Inc.

Hay, Jennifer and Joan Bresnan. 2006. "Spoken syntax: the phonetics of *giving a hand* in New Zealand English", *The Linguistic Review* 23: 321-349.

Hindmarsh, Roland. 1980. *Cambridge English lexicon*. Cambridge: Cambridge University Press.

Hoey, Michael. 2005. *Lexical priming: a new theory of words and language*. Routledge

Huckvale, Mark. 2004. *SCRIBE - Spoken Corpus Recordings In British English*. London: University College London. (http://www.phon.ucl.ac.uk/resource/scribe/scribe-manual.htm) (date of access: 7 Jan. 2012)

Hughes, Arthur and Chryssoula Lascaratou. 1982. "Competing criteria for error gravity", *English Language Teaching Journal* 36, 3: 175-182.

Hulstijn, Jan H. 1992. "Retention of inferred and given word meanings: Experiments in incidental learning", in: Pierre J. Arnaud and Henri Bejoint (eds.), *Vocabulary and applied linguistics*. London: Macmillan, 113-125.

Hulstijn, Jan H. and Batia Laufer. 2001. "Some empirical evidence for the involvement load hypothesis in vocabulary acquisition", *Language Learning* 51, 3: 539–558.

Hyland, Ken and Eri Anan. 2006. "Teachers' perceptions of error: The effects of first language and experience", *System* 34: 509-519.

Hymes, Dell. 1972. "On communicative competence", in: John B. Pride and Janet Holmes (eds.), *Sociolinguistics. Selected readings*. Harmondsworth: Penguin, 269-293.

James, Carl. 1994. "Don't shoot my dodo: On the resilience of contrastive and errors analysis", *International Review of Applied Linguistics in Language Teaching* 32, 3: 179-200.

James, Carl. 1998. *Errors in language learning and use: Exploring error analysis*. London: Longman.

James, Carl. 2005. "Contrastive analysis and the language learner: A new lease of life?" in: David J. Allerton, Cornelia Tshichold, and Judith Wieser (eds.), *Linguistics,*

language learning and language teaching. (The International Cooper Series on English Language and Literature ICSELL 10.) Basel: Schwabe, 1–20.

Jarvis, Scott. 2000. "Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon", *Language Learning* 50, 2: 245-309.

Jarvis, Scott and Aneta Pavlenko. 2008. *Crosslinguistic influence in language and cognition*. New York: Routledge.

Johansson, Stig, Geoffrey N. Leech and Helen Goodluck. 1978. *Manual of information to accompany the Lancaster-Oslo/Bergen Corpus of British English, for use with digital computers.* Oslo: the University of Oslo. (http://khnt.hit.uib.no/icame/manuals/lob) (date of access: 3 Jan. 2012)

Johns, Tim. 1991. "Should you be persuaded: Two examples of data-driven learning", in: Tim Johns and Philip King (eds.), *Classroom concordancing*. Birmingham: English Language Research Journal, 1-13.

Johns. Tim. 1994. "From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning", in: Terence Odlin (ed.), *Perspectives on pedagogical grammar*. Cambridge: Cambridge University Press, 293-313.

Johns, Tim. 1997. Kibbitzers. (http://www.lexically.net/TimJohns/index.html - Mike Scott's personal website) (date of access: 1 May 2012).

Johns, Tim. 2002. "Data-driven learning: the perpetual challenge", in: Bernhard Kettemann and Georg Marko (eds.), *Teaching and learning by doing corpus analysis. Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July, 2000.* Amsterdam: Rodopi, 107-117. (http://www.ingentaconnect.com/content/rodopi/lang/2002/00000042/00000001) (date of access: 20 Sept. 2010)

Johns, Tim, Lee Hsingchin and Wang Lixun. 2008. "Integrating corpus-based CALL programs in teaching English through children's literature", *Computer Assisted Language Learning* 21, 5: 483-506.

Johnson, Keith. 1982. *Communicative syllabus design and methodology*. Oxford: Pergamon.

Kaszubski, Przemysław. 1997. "Polish student writers – Can corpora help them?", in: Barbara Lewandowska-Tomaszczyk and Patrick James Melia (eds.), *PALC'97. International Conference on Practical Applications of Language Corpora. Łódź, Poland, 10-14 April 1997. Proceedings.* Łódź: Łódź University Press, 133- 158.

Kaszubski, Przemysław. 1999. “English learner corpora and the Polish learner”, *Humanising Language Teaching* 1, 8 (http://www.hltmag.co.uk/dec99/idea.htm) (date of access: 7 Jan. 2011).

Kaszubski, Przemysław. 2004. “English learner corpora to the rescue: The case of PICLE”, in: Justyna Leśniewska and Ewa Witalisz (eds.), *The legacy of history. English and American studies and the significance of the past.* Vol. 2: *Language and Culture*. Kraków: Jagellonian University Press, 193-215.

Kaszubski, Przemysław. 2008. “A guided collaboration tool for online concordancing with EFL EAP learners”, in: Ana Frankenberg-Garcia (ed.), *Proceedings of the 8th Teaching and Language Corpora Conference*. Lisbon: Associação de Estudos e de Investigação Cientifica do ISLA–Lisboa, 167-175.

Kaszubski, Przemysław. 2011. “IFAConc – a pedagogic tool for online concordancing with EFL/EAP learners”, in: Ana Frankenberg-Garcia, Lynne Flowerdew and Guy Aston (eds.), *New trends in corpora and language learning*. London: Continuum International Publishing Group, 81-104.

Kellerman, Eric. 1983. “Now you see it, now you don’t”, in: Susan Gass and Larry Selinker (eds.), *Language transfer in language learning*. Rowley, MA: Newbury House, 112-134.

Kellerman, Eric. 1984. “The empirical evidence for the influence of L1 on interlanguage”, in: Alan Davies, Clive Criper and Anthony P. R. Howatt (eds.), *Interlanguage*. Edinburgh: Edinburgh University Press, 98-122.

Kellerman, Eric. 1989. “The imperfect conditional”, in: Kenneth Hyltenstam and Loraine K. Obler (eds.), *Bilingualism across the lifespan: Aspects of acquisition, maturity and loss*. Cambridge: Cambridge University Press, 87-115.

Kim, YouJin. 2008. “The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition”, *Language Learning* 58, 2: 285–325.

Kohn, Kurt. 1986. “The analysis of transfer”, in: Eric Kellerman and Michael Sharwood-Smith (eds.), *Crosslinguistic influence in second language acquisition*. Oxford: Pergamon, 21-34.

Koosha, Masour and Ali Akbar Jafarpour. 2006. “Data-driven learning and teaching collocation of prepositions: the case of Iranian EFL adult learners”, *Asian EFL Journal* 8, 4: 192-209.

Krajka, Jarosław. 2007a. *English language teaching in the Internet-assisted environment: Issues in the use of the web as a teaching medium*. Lublin: Maria Curie-Skłodowska University Press.

Krajka, Jarosław. 2007b. "Corpora and language teachers: From ready-made to teacher-made collections", *CORELL: Computer Resources for Language Learning* 1, 36-55.

Kramsch, Claire. 2007. "Re-reading Robert Lado, 1957, *Linguistics across cultures. Applied linguistics for language teachers*", *International Journal of Applied Linguistics* 17, 2: 241-247.

Krashen, Stephen. 1983. "Newmark's 'ignorance hypothesis' and current second language acquisition theory", in: Susan Gass and Larry Selinker (eds.), *Language transfer in language learning*. Rowley, MA: Newbury House, 135-153.

Krashen, Stephen. 2005. "Free voluntary reading: New research, applications, and controversies", in: Gloria Poediosoedarmo (ed.), *Innovative approaches to reading and writing instruction*, Republic of Singapore: SEAMEO Regional Language Centre. Anthology series 46: 1–9. (http://www.sdkrashen.com/articles/singapore/index.html) (date of access: 7 July 2012)

Krashen, Stephen and Tracy Terrell. 1983. *The natural approach: Language acquisition in the classroom*. Oxford: Pergamon.

Lado, Robert. 1957. *Linguistics across cultures.* University of Michigan Press.

Lamy, Marie-Noëlle and Hans Jørgen Klarskov Mortensen. 2012. "Using concordance programs in the Modern Foreign Languages classroom. Module 2.4", in: Graham Davies (ed.), *Information and communications technology for language teachers (ICT4LT)*. Slough: Thames Valley University (http://www.ict4lt.org/en/en_mod2-4.htm) (date of access: 19 July 2012).

Langacker, Ronald W. 1987. *Foundations of cognitive grammar*. Vol. 1: *Theoretical prerequisites*. Stanford, CA: Stanford University Press.

Larsen-Freeman, Diane. 1986. *Techniques and principles in language teaching*. New York: Oxford University Press.

Larsen-Freeman, Diane. 2003. *Teaching language: from grammar to grammaring*. Toronto: Thomson Heinle.

Larsen-Freeman, Diane and Lynne Cameron. 2008. *Complex Systems and Applied Linguistics.* Oxford: Oxford University Press.

Laufer, Batia and Nany Girsai. 2008. “Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation”, *Applied Linguistics* 29,4: 694-716.

Laurence, Stephen and Eric Margolis. 1999. “Concepts and cognitive science”, in: Eric Margolis and Stephen Laurence (eds.), *Concepts: Core readings*. Cambridge, MA: MIT Press, 3-82.

Lee, Chuen-Yi and Hsien-Chin Liou. 2003. “A study of using web concordancing for English vocabulary learning in a Taiwanese high school context”, *English Teaching and Learning* 27, 3: 35-56.

Leech, Geoffrey. 1994. “Students’ grammar – teachers’ grammar – learners’ grammar”, in: Martin Bygate, Alan Tonkyn and Eddie Williams (eds.), *Grammar and the language teacher*. New York: Prentice Hall, 17-44.

Leech, Geoffrey, Roger Garside and Michael Bryant. 1994. “CLAWS4: The tagging of the British National Corpus”, in: *Proceedings of the 15th International Conference on Computational Linguistics (COLING 94)*. Kyoto, Japan, 622-628. (http://ucrel.lancs.ac.uk/papers/coling1994paper.pdf) (date of access: 10 Feb. 2011).

Lennon, Paul. 1991. “Error: some problems of definition, identification, and distinction”, *Applied Linguistics* 12, 2: 180-196.

Lewandowska-Tomaszczyk, Barbara. 2001. “Computers, language corpora and their applications”, in: Barbara Lewandowska-Tomaszczyk, John Osborne and Frits Schulte (eds.), *Foreign language teaching and information and communication technology*. Frankfurt am Main: Peter Lang, 161-241.

Lewandowska–Tomaszczyk, Barbara, Agnieszka Leńko-Szymańska and Tony McEnery. 2000. “Lexical problem areas in the PELCRA learner corpus of English”, in: Barbara Lewandowska-Tomaszczyk and Patrick James Melia (eds.), *Practial applications in language corpora – Proceedings of PALC ’99*. Frankfurt am Main: Peter Lang, 287-302.

Lewis, Jonathan. 2006. Connecting corpora to learner style: To what extent is the effectiveness of an online corpus-based approach to grammar learning dependent on whether students prefer to learn grammar deductively or inductively? [Unpublished Master’s dissertation. Porto: Universidade do Porto]. (http://www.fc.up.pt/fcup/contactos/teses/t_020370029.pdf) (date of access: 30 Aug. 2012).

Lewis, Michael. 1993. *The lexical approach*. Hove: Language Teaching Publications.

Lewis, Michael. 1997. *Implementing the lexical approach. Putting theory into practice.* Hove: Language Teaching Publications.

Lewis, Michael. 2000. "Language in the lexical approach", in: Michael Lewis (ed.), *Teaching collocation*. Hove: Language Teaching Publications, 126-185.

Liceras, Juana M. 1989. "On some properties of the pro-drop parameter: looking for missing subjects in non-native Spanish", in: Susan Gass and J. Schachter (eds.), *Linguistic perspectives on second language acquisition*. Cambridge: Cambridge University Press, 109-133.

Liski, Erkki and Simo Puntanen. 1983. "A study of the statistical foundations of group conversation tests in spoken English", *Language Learning* 33, 2: 225-246.

Littlewood, William. 1981. *Communicative language teaching*. Cambridge: Cambridge University Press.

Long, Michael H. 1988. "Instructed interlanguage development", in: Leslie M. Beebe (ed.), *Issues in second language acquisition: Multiple perspectives*. New York: Newbury House, 115-141.

Longman dictionary of contemporary English DVD-ROM [*LDOCE5*]. 2009. (5th edition.) Harlow: Pearson Education Limited.

Luka, Barbara and Lawrence W. Barsalou. 1998. "The role of sentence priming on the implicit memoryof syntactic structures", in: Morton Ann Gernsbacher (ed.), *Proceedings of the 20th annual meeting of the Cognitive Science Society*. Mahwah, NJ: Erlbaum. 1240.

Lyons, John. 1968. *An Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.

Lyons, John. 1995. *Linguistic semantics. An introduction.* Cambridge: Cambridge University Press.

Lyster, Roy and Leila Ranta. 1997. "Corrective feedback and learner uptake. Negotiation of form in communicative classrooms", *Studies in Second Language Acquisition* 19, 1: 37-66.

McCarthy, Michael and Felicity O'Dell. 2002. *English vocabulary in use. Advanced*. Cambridge: Cambridge University Press.

McCarthy, Michael and Felicity O'Dell. 2005. *English collocations in use. Intermediate*. Cambridge: Cambridge University Press.

McCarthy, Michael, Jeanne McCarten and Helen Sandiford. 2005. *Touchstone. Student's book 1*. Cambridge: Cambridge University Press.

McCarthy, Michael and Felicity O'Dell. 2008. *Academic vocabulary in use*. Cambridge: Cambridge University Press.

McEnery, Tony and Andrew Wilson. 1996. *Corpus linguistics*. Edinburgh: Edinburgh University Press.

Manning, Christopher D. and Hinrich Schütze. 2001. *Foundations of statistical natural language processing*. Cambridge, MA: The MIT Press.

Marcus, Gary. 2009. "How *does* the mind work? Insights from biology", *Topics in Cognitive Science* 1: 145-172.

Mattar, Hameed. 1999. "Translation elicitation techniques and mother-tongue interference: Any Significant Connection?", *International Review of Applied Linguistics* 37, 4: 307-320.

Meriläinen, Lea. 2010. "Syntactic transfer in the written English of Finnish students: Persistent grammar errors or acceptable lingua franca English?", *Apples – Journal of Applied Language Studies* 4, 1: 51-64.

Meriö, Katri. 1978. "The psycholinguistic analysis and measurement of interference errors", *International Review of Applied Linguistics* 16, 1: 27-44.

Mezynski, Karen. 1983. "Issues concerning the acquisition of knowledge: Effects of vocabulary training on reading comprehension", *Review of Educational Research* 53, 2: 253-279.

Mougeon, Raymond and Pierre Hébrard. 1975. *L'acquisition et la maîtrise de l'anglais par les jeunes bilingues de Welland*. Toronto: Ontario Institute for Studies in Education.

Murphy, Gregory L. 2004. *The big book of concepts*. Cambridge, MA: MIT Press.

Norris, John M. and Lourdes Ortega. 2000. "Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis", *Language Learning* 50, 3: 417-528.

Odlin, Terence. 1989. *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.

Odlin, Terence. 1994. *Perspectives on pedagogical grammar*. Cambridge: Cambridge University Press.

O'Keefe, Anne and Fiona Farr. 2003. "Using language corpora in initial teacher education: Pedagogic issues and practical applications", *TESOL Quarterly* 37, 3: 389-418.

O'Keeffe, Anne, Michael McCarthy and Ronald Carter. 2007. *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.

Oller John and Elcho Redding. 1971. "Article usage and other language skills", *Language Learning* 21, 1: 85-95.

Osborne, John. 2004. "Top-down and bottom-up approaches to corpora in language teaching", in: Ulla Connor and Thomas A. Upton (eds.), *Applied corpus linguistics. A multidimensional perspective.* (Language and Computers 52.) Amsterdam: Rodopi. 251-265. (http://www.ingentaconnect.com/content/rodopi/ lang/ 2004/00000052/00000001) (date of access 26 Nov. 2010).

Paribakht, T. Sima and Marjorie Wesche. 1997. "Vocabulary enhancement activities and reading for meaning", in: James Coady and Thomas Huckin (eds.), *Second language vocabulary acquisition*. Cambridge: Cambridge University Press, 174-200.

Partington, Alan. 1998. *Patterns and meanings: Using corpora for English language research and teaching*. Amsterdam: John Benjamins Publishing Company.

Pawlak, Mirosław. 2006. *The place of form-focused instruction in the foreign language classroom*. Kalisz: Uniwersytet im. Adama Mickiewicza.

Pawlak, Mirosław. 2010. "Learners' perspective on the role of error correction in the foreign language classroom: The importance of space, techniques and timing", in: Andrzej Ciuk and Katarzyna Molek-Kozakowska (eds.), *Exploring space: Spacial notions in cultural, literary and language studies*. Vol. 2. Newcastle upon Tyne: Cambridge Scholars Publishing, 316-325.

Pawlak Mirosław. 2012. *Error correction in the foreign language classroom: Reconsidering the issues*. Poznań: Uniwersytet im. Adama Mickiewicza.

Pawley, Andrew and Frances Hodgetts Syder. 1983. "Two puzzles for linguistic theory: Nativelike selection and nativelike fluency", in: Jack C. Richards and Richard W. Schmidt (eds.), *Language and Communication*. London: Longman, 191-226.

Peacock, Matthew. 1999. "Beliefs about language learning and their relationship to proficiency", *International Journal of Applied Linguistics* 9, 2: 247-265.

Phillips, Denis Charles. 2000. “An opinionated account of the constructivist landscape”, in: Denis Charles Phillips (ed.), *Constructivism in education. Opinions and second opinions on controversial issues*. Chicago, IL: The University of Chicago Press, 1-16.

Piaget, Jean. 1980. “The psychogenesis of knowledge and its epistemological significance”, in: Massimo Piattelli-Palmarini (ed.), *Language and learning. The debate between Jean Piaget and Noam Chomsky*. Cambridge, MA.: Harvard University Press.

Pinker, Steven and Alan Price. 1988. “On language and connectionism: Analysis of a parallel distributed processing model of language acquisition”, *Cognition* 28: 73-193.

Popper, Karl. 1968. “Epistemology without a knowing subject”, *Studies in Logic and the Foundations of Mathematics* 52, 333–373. (http://scholar.googleuser content.com/scholar?q=cache:kyLkmXG5QMcJ:scholar.google.com/&hl=en&as_sdt=1,5) (date of access: 23 Feb. 2012).

Prince, Alan and Paul Smolensky. 2004. *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell Publishing.

Proulx, Jérôme. 2006. “Constructivism: A re-equilibration and clarification of the concepts, and some potential implications for teaching and pedagogy”, *Radical Pedagogy* 8, 1. (radicalpedagogy.icaap.org/content/issue8_1/proulx.html) (date of access: 23 Feb. 2012)

Purcell, Edward and Richard Suter. 1980. “Predictors of pronunciation accuracy”, *Language Learning* 30: 271-287.

Ragan, Peter H. 2001. “Classroom use of a systemic functional learner corpus”, in: Mohsen Ghadessy, Alex Henry, Robert L. Roseberry, *Small corpus studies and ELT: Theory and practice*. Amsterdam: John Benjamins Publishing Company.

Richards, Jack. 1971. “A non-contrastive approach to error analysis”, *ELT Journal* 25: 204-219.

Richards, Jack and Theodore Rogers. 2001. *Approaches and methods in language teaching*. (2nd edition.) Cambridge: Cambridge University Press.

Roberts, Michael and Carol Griffiths. 2008. “Errors and good language learners” in: Carol Griffiths (ed.), *Lessons from good language learners*. Cambridge: Cambridge University Press, 282-293.

Robinson, Peter. 1995. "Attention, memory and the 'noticing hypothesis", *Language Learning* 45, 2: 283-331.

Rogers, Margaret. 1984. "On major types of written error in advanced students of German", *International Review of Applied Linguistics* 22, 1: 1-39.

Römer, Ute. 2008. "Corpora and language teaching", in: Anke Lüdeling and Merja Kytö (eds.), *Corpus linguistics : An international handbook*. Berlin: Walter de Gruyter, 112-131.

Römer, Ute. 2011. "Corpus research applications in second language teaching", *Annual Review of Applied Linguistics* 31: 205-2005.

Rosch, Eleanor. 1975. "Cognitive representations of semantic categories", *Journal of Experimental Psychology: General* 104: 192-233.

Rosch, Eleanor and Carolyn B. Mervis. 1975. "Family resemblances: Studies in the internal structure of categories", *Cognitive Psychology* 7: 573-605.

Rumelhart, David E. and James L. McClelland. 1987. "Learning the past tenses of English verbs: Implicit rules or parallel distributed processing?", in: Brian MacWhinney (ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum, 194-248.

Rundell, Michael (ed.). 2007. *Macmillan English dictionary for advanced learners* [*MED2*]. (2nd edition.) Oxford: Macmillan Education.

Rundell, Michael and Sylviane Granger. 2007. "From corpora to confidence", *English Teaching Professional* 50: 15-18. (http://www.macmillandictionaries.com/reviews/documents/Fromcorporatoconfidence-EnglishTeachingProfessional-May2007.pdf). (date of access: 7 Feb. 2011).

St.John, Elke. 2001. "A case for using a parallel corpus and concordancer for beginners of a foreign language", *Language Learning and Technology* 5, 3: 185-203.

Salem, Ilana. 2007. "The lexico-grammatical continuum viewed through student error", *ELT Journal* 61, 3: 211-219.

Saxton, Matthew. 1997. "The contrast theory of negative input", *Journal of Child Language* 24, 1: 139-161.

Sapir, Edward. 1921. *Language: An introduction to the study of speech.* New York: Harcourt, Brace & World.

Schachter, Jacquelyn, Adele F. Tyson and Frank J. Diffley. 1976. "Learner intuitions of grammatically", *Language Learning* 26, 1: 67-76.

Schachter, Jacquelyn and Marianne Celce Murcia. 1977. “Some reservations concerning error analysis”, *TESOL Quarterly* 11: 441-451.

Scheffler, Paweł. 2008. *Native English and learner corpora: Linguistic comparison and pedagogical applications.* Poznań: Wydawnictwo Naukowe UAM.

Schmidt, Richard A. and Robert A. Bjork. 1992. “New conceptualizations of practice: common principles in three paradigms suggest new concepts for training”, *Psychological Science* 3, 207-217.

Schmidt, Robert. 1990. “The role of consciousness in second language learning”, *Applied Linguistics* 11: 129-158.

Scholfield, Phil. 1995. *Quantifying language: A researcher's and teacher's guide to gathering language data and reducing it to figures*. Clevendon: Multilingual Matters.

Schumann, John H. and Nancy Stenson (eds.). 1976. *New frontiers in second language learning*. Rowley, Mass.: Newbury House.

Scott, Margaret Sue and G. Richard Tucker. 1974. “Error analysis and English-language stra-tegies of Arab students”, *Language Learning* 24: 69-97.

Seliger, Herbert, Stephen Krashen and Peter Ladefoged. 1975. “Maturational constraints in the acquisition of second languages”, *Language Sciences* 38: 20-22.

Selinker, Larry. 1972. “Interlanguage”, *International Review of Applied Linguistics* 10, 3: 209-231.

Selinker, Larry. 1992. *Rediscovering interlanguage*. London: Longman.

Selinker, Larry, Merrill Swain and Guy Dumas. 1975. “The interlanguage hypothesis extended to chidren”, *Language Learning* 25: 139-152.

Selinker, Larry and Usha Lakshmanan. 1992. “Language transfer and fossilization; the Multiple Effects Principle”, in: Susan M. Gass and Larry Selinker (eds.), *Language transfer in language learning*. Amsterdam: John Benjamins Publishing Company,197-216.

Sheen, Ronald. 1996. “The advantage of exploiting contrastive analysis in teaching and learning a foreign language”, *International Review of Applied Linguistics* 34, 3: 183-197.

Siek-Piskozub, Teresa. 2006. “Constructivism in language pedagogy”, in: Edyta Lorek-Jezińska, Teresa Siek-Piskozub and Katarzyna Więckowska (eds.), *Worlds in the*

making. Constructivism and postmodern knowledge. Toruń: Wydawnictwo Uniwersytetu Mikołaja Kopernika, 159-172.

Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, John. 1998. "The lexical item", in: Edda Weigand (ed.), *Contrastive lexical semantics*, Amsterdam: John Benjamins, 1-24.

Sinclair, John (ed.). 1987. *Collins COBUILD English language dictionary.* London: Collins.

Sinclair, John (ed.). 1990. *Collins COBUILD English grammar*. London: HarperCollins Publishers.

Sinclair, John and Malcolm Coulthard. 1975. *Towards an analysis of discourse*. Oxford: Oxford University Press.

Skinner, Burrhus Frederic. 1957. *Verbal behavior*. New York: Appleton-Century-Crofts.

Soloman, Barbara A. and Richard M. Felder. 2012. *Index of Learning Styles Questionnaire* [*ILSQ*]. North Carolina State University. (http://www.engr.ncsu.edu/learningstyles/ilsweb.html) (date of access: 30 Aug. 2012).

Spitzer, Manfred. 1999. *The mind within the net*. Cambridge, MA: MIT Press.

Spitzer, Manfred. 2006. "Brain research and learning over the life cycle", in: *Introduction to personalisation*. Paris: OECD Schooling for Tomorrow. (http://www.oecd.org/dataoecd/51/38/41176708.pdf) (date of access: 23 April 2012).

Sripicharn, Passapong. 2003. "Evaluating classroom concordancing: The use of concordance-based materials by a group of Thai students", *Thammasat Review* 8, 1: 203-236. (http://www.tu.ac.th/resource/publish/interview/vol8.html) (date of access: 20 Jun. 2012).

Stemberger, Joseph Paul. 1982. "Syntactic errors in speech", *Journal of Psycholinguistic Research* 11, 4: 313-345.

Stenson, Nancy. 1974. "Induced errors", in: John H. Schumann and Nancy Stenson (eds.), *New frontiers in second language learning*. Rowley, MA: Newbury House, 54-70.

Stern, Hans Heinrich. 1992. *Issues and options in language teaching*. Oxford: Oxford University Press.

Sternberg, Robert J. 1987. “Most vocabulary is learned from context”, in: Margaret G. McKeown and Mary E. Curtis (eds.), *The nature of vocabulary acquisition*. Hillsdale, NJ: Erlbaum, 89-105.

Stockwell, Robert, J. Donald Bowen and John W. Martin. 1965. *The grammatical structures of English and Spanish: An analysis of structural differences between the two languages*. Chicago: The University of Chicago Press.

Strapp, Chehalis M., Augusta L. Helmick, Hayley M. Tonkovich and Dana M. Bleakney. 2011. “Effects of negative and positive evidence on adult word learning”, *Language Learning* 62, 2: 506-532.

Strunk, William, Jr. and E.B. White. [1918] 2009. *The elements of style*. (5th edition). Boston: Allyn and Bacon.

Summers, Della (ed.). 1993. *Longman language activator*. Harlow: Longman.

Summers, Della (ed.). 1995. *Longman dictionary of contemporary English*. Harlow: Longman.

Summers, Della (ed.). 1997. *Longman essential activator*. Harlow: Longman.

Sun, Yu-Chih, and Li-Yuch Wang. 2003. “Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty”, *Computer Assisted Language Leaning* 16, 1: 83-94.

Swan, Michael. 1995. *Practical English usage*. (2nd edition.) Oxford: Oxford University Press.

Swan, Michael. 2005. *Practical English usage*. (3rd edition.) Oxford: Oxford University Press.

Swan, Michael. 2007. “Follow-up to Claire Kramsch’s ‘classic book’ review of Lado, 1957, *Linguistics across cultures*: History is not what happened: the case of contrastive analysis”, *International Journal of Applied Linguistics* 17, 3: 414-419.

Szulc, Aleksander. 1984. *Podręczny słownik językoznawstwa stosowanego*. Warszawa: Państwowe Wydawnictwo Naukowe.

Tarone, Elaine. 1979. “Interlanguage as chameleon”, *Language Learning* 29, 181-191.

Tarone, Elaine. 1983. “On the variability of interlanguage systems”, *Applied Linguistics* 4, 142-163.

Tarone, Elaine. 1985. “Variability in interlanguage use: A study of style-shifting in morphology and syntax”, *Language Learning* 35, 373-403.

Taylor, Barry P. 1975. “The use of overgeneralization and transfer strategies by elementary and intermediate students of ESL”, *Language Learning* 25, 1: 73-107.

Thomason, Sarah and Terrence Kaufmann. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Tomasello, Michael and Carol Herron. 1989. “Feedback for language transfer errors: The garden path technique”, *Studies in Second Language Acquisition* 11, 4: 385-395.

Tomlin, Russel S. and Victor Villa. 1994. “Attention in cognitive science and second language acquisition”, *Studies in Second Language Acquisition* 16, 2: 183-203.

Turnbull, Jill and Jack Burston. 1998. “Towards independent concordance work for students: Lessons from a case study”, *ON-CALL*, 12,2: 10-21.

Turton, Nigel and Brian Heaton. 1996. *Longman dictionary of common errors*. Harlow: Longman.

Upton, Thomas and Ulla Connor. 2001. “Using computerized corpus analysis to investigate the textlinguistic discourse moves of a genre”, *English for Specific Purposes* 20: 313-329.

von Glasersfeld, Ernst. 1997. “Homage to Jean Piaget”, *Irish Journal of Psychology* 18, 2: 293-306.

Weinreich, Uriel. [1953] 1968. *Languages in contact: Findings and problems.* The Hague: Mouton. [Originally published, New York: Linguistic Circle of New York.].

White, Lydia. 1985. “The pro-drop parameter in adult second language acquisition”, *Language Learning* 35: 47-62.

White, Lydia. 1992. “Universal Grammar: Is it just a new name for old problems?”, in: Susan Gass and Larry Selinker (eds.), *Language transfer in language learning*. (revised edition) Amsterdam: John Benjamins Publishing Company, 217-236. ?

White, Lydia. 2003. *Second language acquisition and universal grammar*. Cambridge: Cambridge University Press.

Widdowson, Henry George. 1972. “The teaching of English as communication”, *English Language Teaching* 27, 1: 15-18.

Widdowson, Henry George. 1979. *Explorations in applied linguistics.* Oxford: Oxford University Press.

Willim, Ewa and Elżbieta Mańczak-Wohlfeld. 1997. *A contrastive approach to problems with English*. Warszawa: Wydawnictwo Naukowe PWN.

Willis, Dave and Jane Willis. 1989. *Collins COBUILD English Course* [*CCEC*]. London: HarperCollins.

Wolański, Adam Filip. 2008. *Edycja tekstów: praktyczny poradnik : książka, prasa, www* [Text editing: A practical guide: Books, press, www]. Warszawa: Wydawnictwo Naukowe PWN.

Woolard, George. 2000. "Collocation – encouraging learner independence", in: Michael Lewis (ed.), *Teaching collocation. Further developments in the lexical approach*. Hove: Language Teaching Publications, 28-46.

Yip, Virginia. 1994. "Grammatical consciousness raising and learnability", in: Terence Odlin (ed.), *Perspectives on pedagogical grammar*. Cambridge: Cambridge University Press, 123-138.

Corpora and text analysis tools

Anthony, Laurence. 2012. *AntWordProfiler*. (http://www.antlab.sci.waseda.ac.jp/antwordprofiler_index.html) (date of access: 20 Nov. 2012).

BNCweb. 2008. Lancaster: Lancaster University. (http://bncweb.lancs.ac.uk) (date of access: 28 Apr. 2012).

Cobb, Tom. 2011. *KeyWords*. Edict. (http://www.edict.biz/pub/concapp/) (date of access: 9 Feb. 2011).

Davies, Mark. 2004-. *BYU-BNC*. (Based on the British National Corpus from Oxford University Press). (http://corpus.byu.edu/bnc/) (date of access 26 July 2012).

Davies, Mark. 2008-. *The Corpus of Contemporary American English: 450 million words, 1990-present*. (http://corpus.byu.edu/coca/) (date of access 26 July 2012).

Edict. 2011. *ConCapp* (http://www.edict.biz/pub/concapp/) (date of access: 9 Feb. 2011).

Edict. 2011. *Word Frequency Text Profiler*. (http://www.edict.biz/textanalyser/) (date of access: 9 Feb. 2011).

Hrdina, Magdalena, Kamil Sobczak, and Marcin Sochacki. (2012) *LAGUN Corpus*. (http://korpus.hiztegia.org) (date of access: 3 Aug. 2012).

Hüning, Matthias. 2008. *TextSTAT*. (http://www.niederlandistik.fu-berlin.de/textstat/) (date of access: 22 July 2009).

Kaszubski, Przemysław. 2002. *PICLE: the Polish sub-corpus of the International Corpus of Learner English* (http://ifa.amu.edu.pl/~ifaconc/blog/?page_id=60) (date of access: 3 Jan. 2013).

Pęzik, Piotr. 2012. *Wyszukiwarka PELCRA dla danych NKJP* [PELCRA search engine for the NCP]. Adam Przepiórkowski, Mirosław Bańko, Rafał Górski and Barbara Lewandowska-Tomaszczyk (eds.), *Narodowy Korpus Języka Polskiego* [The National Corpus of Polish]. Warszawa: Wydawnictwo Naukowe PWN. (http://www.nkjp. uni.lodz.pl/) (date of access: 26 Sept. 2012).

Scott, Mike. 2012. *WordSmith Tools*. Lexical Analysis Software Ltd. (product support website) (http://www.lexically.net/wordsmith/index.html) (date of access: 9 Feb. 2011).

Statistical analysis software:

IBM SPSS Statistics 19. 2010. IBM Corporation.

Microsoft Office Excel. 2007. Microsoft Corporation.

Appendix A: Examination errors database

No.	Student Code	error	Type	Sub-type	Domain	Surface strategy	(Language problem) > plausible reconstruction	Plausible explanation and associated L1 feature
1	2034	to communicate with OTHER and...	form	categorial	phrase	misselection	pro-form > OTHERS (determiner vs. pronoun)	underdifferentiation - PL: inny/inni>other/the other/another/others/the others
2	2015	his or HERS extraordinary photo	form	categorial	phrase	misselection	pro-form > his or HER extraordinary photo (determiner vs. pronoun)	underdifferentiation - PL: no difference in form (possessive det/pron)
3	2010	loom over the HORIZONT	form	grapho-phon	word	overinclusion	> horizon	foreignizing: horyzont (formal similarity)
4	2020	an "earlier start" even ENCHANCES the child's ... abilities	form	grapho-phon	clause	overinclusion	> enhances	underdifferentiation: PL - no phonetic differentiation between [h]/[ch]
5	2026	papparrazi	form	grapho-phon	word	overinclusion/ omission	> paparazzi; single cons. letter doubled, double cons - reduced to one	L1 feature absent from L2: PL - gemination occurs if consonant letters doubled
6	2015	Pulizer's (sic!) PRICE	form	grapho-phon	phrase	misselection	> prize	underdifferentiation: PL - no /s/ vs. /z/ contrast in word-final position (final devoicing)
7	2026	as a role MODAL	form	grapho-phon	phrase	misselection	> model	feature absent from L1: no vowel reduction in PL
8	2010	enlarge (sic!) the STUFF number	form	grapho-phon	phrase	misselection	> increase the number of STAFF	underdifferentiation: PL - only one vowel in the low-back region: problems with /ʌ/vs./æ/ distinction -
9	2007	careless EXISTANCE	form	grapho-phon	word	misselection	> existence	feature absent from L1: sound-to-spelling problems: /ə/ - no corresponding vowel in PL
10	2026	a role MODAL	form	grapho-phon	phrase	misselection	> model	feature absent from L1: sound-to-spelling problems: /ə/ - no corresponding vowel in PL
11	2015	VERTUES	form	grapho-phon	word	misselection	> (virtues)	feature absent from L1: sound-to-spelling problems: /ɜ/&/ɝ/ - no corresponding vowel in PL, /e/ is chosen as the closest one
12	2038	can be EASLY noticed	form	grapho-phon	word	omission	> easily	feature absent from L1: inaudible (reduced) vowel omitted in spelling
13	2007	they cannot AFORD it	form	grapho-phon	word	omission	> afford	L1 feature absent from L2: PL - gemination occurs if consonant letters doubled
14	2015	chalenging	form	grapho-phon	word	omission	> challenging	L1 feature absent from L2: PL - gemination occurs if consonant letters doubled
15	2013	Since it is GENERALY known	form	grapho-phon	word	omission	> generally	L1 feature absent from L2: PL - gemination occurs if consonant letters doubled
16	2015	wrong and INAPROPRIATE	form	grapho-phon	word	omission	> inappropriate	L1 feature absent from L2: PL - gemination occurs if consonant letters doubled
17	2020	future INTELECTUAL abilities	form	grapho-phon	word	omission	> INTELLECTUAL	L1 feature absent from L2: PL - gemination occurs if consonant letters doubled
18	2029	to show something only PARTIALY	form	grapho-phon	word	omission	> partially	L1 feature absent from L2: PL - gemination occurs if consonant letters doubled
19	2029	it is PARTIALY true	form	grapho-phon	word	omission	> partially true	L1 feature absent from L2: PL - gemination occurs if consonant letters doubled

20	2031	having their tragedies shown publicaly	form	grapho-phon	word	omission	> publically	L1 feature absent from L2: PL - gemination occurs if consonant letters doubled
21	2034	successfuly	form	grapho-phon	word	omission	> successfully	L1 feature absent from L2: PL - gemination occurs if consonant letters doubled
22	2013	I TOTALY agree	form	grapho-phon	word	omission	> totally	L1 feature absent from L2: PL - gemination occurs if consonant letters doubled
23	2039	a constant and strong WILLINGNES to study	form	grapho-phon	word	omission	> willingness	L1 feature absent from L2: PL - gemination occurs if consonant letters doubled
24	2004	GOVERMENT officials	form	grapho-phon	word	omission	> government officials	feature absent from L1: PL - no silent letters other than in digraphs
25	2019	CLONS	form	grapho-phon	word	omission	> clones	feature absent from L1: silent vowel letter - not used in PL
26	2029	it is IMMENSLY difficult	form	grapho-phon	word	omission	> immensely	feature absent from L1: silent vowel letter - not used in PL
27	2014	8 golden (sic!) medals	form	grapho-phon	text	misselection	> EIGHT gold medals	feature absent from L1: English formal style restriction re numbers not used in PL
28	2010	the first 6 or 7 years	form	grapho-phon	text	misselection	> the first SIX or SEVEN years	feature absent from L1: English formal style restriction re numbers not used in PL
29	2014	winning 8 golden medals	form	grapho-phon	text	misselection	> winning EIGHT golden medals	feature absent from L1: English formal style restriction re numbers not used in PL
30	2003	six years old children	form	infl	phrase	overinclusion	sing/pl noun modifier - > six-year-old children	feature absent from L1: sześcioletnie dzieci (Num + sing noun not used in PL)
31	2015	Pulizer'S price (sic!)	form	infl	phrase	misselection	genitive vs. compound noun > Pulit-zer Prize	calque: PL: Nagroda Pulitzera (PL: genitive)
32	2004	there is every evidence that the VIEW OF SOME PARENTS that…	form	infl	clause	misselection	Sax gen - plural > some parents' view that...	underdifferentiation: PL - one genitive form for human and non-human nouns (pogląd niektórych rodziców, że...)
33	2033	shaping the minds OF PEOPLE	form	infl	phrase	misselection	Sax gen vs. of-gen > shaping people's minds?	underdifferentiation: PL - one genitive form for human and non-human nouns (umysły ludzi)
34	2008	the fact that they are creating opinion and view of many people	form	infl	phrase	misselection	Sax gen vs. of-gen > that they are forming MANY PEOPLE'S opinions and views	
35	2010	the majority of the MATURE'S memories of their childhood IS closely related…	form	infl	phrase	misselection	Sax gen vs. of-gen > the majority of MATURE people's memories of their childhood are closely related…	underdifferentiation: PL genitive - inflectional for all nominal items; here - not possible (gener-ic adjective, not 'real' noun, cannot be used with Sax gen)
36	2026	a high class SWIMMERS poor behaviour	form	infl	phrase	omission	apostrophy > a high class SWIM-MER'S poor behaviour	feature absent from L1: punctuation not used for marking inflection in PL
37	2026	in TODAYS society	form	infl	phrase	omission	apostrophy: Sax genitive > in TO-DAY'S society	feature absent from L1: punctuation not used for marking inflection in PL
38	2034	PARENTS attitude towards their children	form	infl	phrase	omission	apostrophy: Sax genitive - plural > parents' attitude	feature absent from L1: punctuation not used for marking inflection in PL
39	2034	ruin their CHILDREN future	form	infl	phrase	omission	Sax gen - plural > children's future	feature absent from L1: punctuation not used for marking inflection in PL

40	2034	for their CHILDREN sake	form	infl	phrase	omission	Sax gen - plural > for their children's sake	feature absent from L1: punctuation not used for marking inflection in PL
41	2038	about the children' abilities in the learning process	form	infl	phrase	omission	Sax gen - plural > the children's abilities...	feature absent from L1: punctuation not used for marking inflection in PL
42	2005	CHILDREN needs	form	infl	phrase	omission	Sax gen > children's	feature absent from L1: punctuation not used for marking inflection in PL
43	2016	the US Olympic Committee future actions	form	infl	phrase	omission	Sax gen > the US Olympic Committee's future actions	feature absent from L1: punctuation not used for marking inflection in PL
44	2016	the recent British newspaper PUBLISHMENT	form	wf	word	misselection	> publication	coinage: overgeneralization of a deriv. suffix - publikacja (kura-cja > treat-MENT, inwesty-cja>invest-MENT, etc.)
45	2038	a proper amount of DOTATIONS	form	wf	word	misselection	(non-existent word!) > subsidies?	foreignizing: dotacje
46	2038	PSYCHOLOGIC in nature	form	wf	word	misselection	> psychological	foreignizing: natury psychologicznej (formal similarity)
47	2020	psychologic	form	wf	word	misselection	> psychological	foreignizing: psychologiczna (formal similarity)
48	2002	mathematic exercises	form	wf	word	misselection	> mathematical exercises/maths problems	foreignizing: matematyczne (formal similarity)
49	2014	coverage of anyone of them	form	wf	phrase	misselection	> coverage of any (one) of them	underdifferentiation - ktokolwiek (anyone/any of them/any one of them - emphatic): materiał o którymkolwiek z nich
50	2005	real money that ENTER national budget	lex-gram	agr	clause	misselection	> enters? is put in the budget?	calque: (money) PL pieniądze - plural
51	2005	money RULE the world	lex-gram	agr	clause	misselection	> money RULES the world	calque: (money) PL pieniądze - plural
52	2025	news which people think to be objective SOAK with political ... lobbying	lex-gram	agr	clause	misselection	news - sing > IS saturated with (?)	calque: (news)PL wiadomości - plural
53	2029	the INFORMATION that sometimes ARE not worth seeing	lex-gram	agr	clause	misselection	> the INFORMATION that sometimes IS not worth seeing	calque: information - plural in Polish > plural verb used (induced error also possible)
54	2031	the most problematic media's target ARE the victims of tragedies	lex-gram	agr	clause	misselection	> media's most problematic target IS the victims...	calque: najbardziej problematycznym odbiorcą mediów są ofiary (Polish concord different)
55	2002	students of various age	lex-gram	agr	phrase	misselection	> students of various ages	calque: studenci w różnym wieku
56	2002	reading and writing SKILL	lex-gram	agr	phrase	misselection	> reading and writing skills	calque: umiejętność czytania i pisania - singular concept in Polish
57	2014	The vast majority [of students - A.L.] QUITS drugs…	lex-gram	agr	clause	misselection	> quit	calque: Większość (studentów) RZUCA narkotyki (different concord rules in Polish) - "większość" not a modifier, but a head noun
58	2010	the majority of the mature's (sic!) memories of their childhood IS closely related…	lex-gram	agr	clause	misselection	> the majority of ... memories ARE closely related...	calque: większość wspomnień… JEST
59	2015	the media ITSELF often MAKE new definition	lex-gram	agr	clause	misselection	> media themselves OR > media ... makes (blend?)	feature absent from L1: reflexive emphatic pronoun error - a much more complex system than in Polish, where one lexical item is inflected for gender and number

60	2025	one can use their common sense	lex-gram	agr	clause	misselection	> one can use one's common sense	feature absent from L1: reflexive possessive pronoun error - a much more complex system than in Polish, where one lexical item (swój) is inflected for gender and number
61	2027	schools itself are not equipped...	lex-gram	agr	phrase	misselection	reflexive pron > schools themselves are not equipped...	feature absent from L1: reflexive pronoun error - a much more complex system than in Polish, where one lexical item is inflected for gender and number - same szkoły nie są wyposażone... ("archi-form" ITSELF?)
62	2038	the number of births DECREASE every year	lex-gram	agr	clause	misselection	> decreases	transfer of learning - 'number' is taught as a problematic word concerning agreement, which may have caused the confusion
63	2002	as * result	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
64	2003	there are * number of	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
65	2003	build up * economic system	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
66	2003	* better chance to stay in business	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
67	2003	* seven year old child	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
68	2003	is rather * combination of what	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
69	2003	to have * balanced view	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
70	2005	whether to send * child to school	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
71	2005	If * child is not ready…	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
72	2007	the most significant part of * human's life	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
73	2010	the first 6 or 7 years of * child's life	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
74	2010	for * child to take up	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
75	2013	such * splendid career	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
76	2014	which is nothing but * burden	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
77	2019	achieve * really startling level	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
78	2027	to find * better job in future	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
79	2029	* rather subjective perspective	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
80	2029	there is also * second vastly im-portant point	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
81	2032	the idea of * happy childhood	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
82	2034	in (sic!) * young age	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
83	2034	has * significant effect	lex-gram	art	phrase	omission	> "a"	feature absent from L1: no articles in PL
84	2004	require * enormous amount	lex-gram	art	phrase	omission	> "a" OR plural> enormous amounts	feature absent from L1: no articles in PL
85	2005	invest * huge amount of money	lex-gram	art	phrase	omission	> "a" OR plural> huge amounts	feature absent from L1: no articles in PL
86	2005	* Economical crissis (sic!)	lex-gram	art	phrase	omission	> "an"	feature absent from L1: no articles in PL
87	2015	* Afro-american (sic!) becomes the President…	lex-gram	art	phrase	omission	> "an"	feature absent from L1: no articles in PL
88	2025	* Important issue is that...	lex-gram	art	phrase	omission	> "An"	feature absent from L1: no articles in PL
89	2029	* Outstandingly high percentage of well educated, smart citizens	lex-gram	art	phrase	omission	> "An"	feature absent from L1: no articles in PL

90	2034	A child should have * opportunity to...	lex-gram	art	phrase	omission	> "an"	feature absent from L1: no articles in PL
91	2034	the most important in one's future is * ability to communicate...	lex-gram	art	phrase	omission	> "an"	feature absent from L1: no articles in PL
92	2005	* national budget	lex-gram	art	phrase	omission	> "the"	feature absent from L1: no articles in PL
93	2005	on the basis of * child's … abilities	lex-gram	art	phrase	omission	> "the"	feature absent from L1: no articles in PL
94	2005	also * child will surely not reach	lex-gram	art	phrase	omission	> "the"	feature absent from L1: no articles in PL
95	2005	if * child is ready	lex-gram	art	phrase	omission	> "the"	feature absent from L1: no articles in PL
96	2015	putting a beautiful model in * star's way	lex-gram	art	phrase	omission	> "the"	feature absent from L1: no articles in PL
97	2016	escaping from * problem	lex-gram	art	phrase	omission	> "the"	feature absent from L1: no articles in PL
98	2025	* Chase for sensational materials	lex-gram	art	phrase	omission	> "the"	feature absent from L1: no articles in PL
99	2025	* incredible speed of information	lex-gram	art	phrase	omission	> "the"	feature absent from L1: no articles in PL
100	2029	we should focus on * professional side of media	lex-gram	art	phrase	omission	> "the"	feature absent from L1: no articles in PL
101	2034	the neutrality of a place and * teacher	lex-gram	art	phrase	omission	> "the"	feature absent from L1: no articles in PL
102	2034	problems of * outside world	lex-gram	art	phrase	omission	> "the"	feature absent from L1: no articles in PL
103	2038	the history of * Polish educational system	lex-gram	art	phrase	omission	> "the"	feature absent from L1: no articles in PL
104	2038	that * Polish educational system...	lex-gram	art	phrase	omission	> "the"	feature absent from L1: no articles in PL
105	2036	with a little amount of * sport superstars' willingness...	lex-gram	art	phrase	omission	> "the" (context: singular definite reference)	feature absent from L1: no articles in PL
106	2005	to begin voyage into * unknown	lex-gram	art	phrase	omission	> into THE unknown	feature absent from L1: no articles in PL with generic adjectives
107	2005	enter * job market	lex-gram	art	phrase	omission	> "the" > enter THE job market	feature absent from L1: wejdą na rynek pracy
108	2008	one of January issues of SOME British newspaper	lex-gram	art	phrase	misselection	determiner (negative semantic prosody) > A British newspaper	underdifferentiation - jakiś (absence of indefinite articles in PL enhamces the use of the more MARKED form "some")
109	2038	kindergartens will have to FACE WITH the extraordinary number of...	lex-gram	colligation	phrase	overinclusion	prep insertion > kindergartens will have to face Ø the extraordinary number of...	calque: będą musiały stanąc twarzą w twarz Z (czymś)… /zmierzyć się Z czymś
110	2027	the difficulties the Kindergartens have to tackle WITH (sth)...	lex-gram	colligation	phrase	overinclusion	> the difficulties the Kindergartens have, tackling Ø (sth)... (no preposition)	calque: radzić sobie Z czymś
111	2032	younger learners are SUSCEPTIBLE TO MISUNDERSTAND ... instructions	lex-gram	colligation	clause	blend	ing/inf > are susceptible/prone to + N/-ing OR TEND to misunderstand	calque: skłonni sa mylnie rozumieć instrukcje
112	2019	such a ... superstar LIKE M.P.	lex-gram	colligation	phrase	blend	comparative phrase structure > such a superstar as...	underdifferentiation - like/as - PL "jak"
113	2036	IT IS without a shadow of a doubt to me that...	lex-gram	colligation	sentence	blend	"empty" subject > There is no doubt that... (OR: I KNOW without a shadow of a doubt)	underdifferentiation: an impersonal form - Nie ma dla mnie cienia wątpliwości, że..

114	2038	people will decide ON STARTING the (sic!) family life	lex-gram	colligation	phrase	misselection	ing/inf > to start ("decide" + gerund - 3 scores in the BNC; "decide" + inf - 1064 scores)	calque: ludzie zdecyduja się na rozpoczęcia życia rodzinnego
115	2033	Those who are vulnerable to be affected by this fourth power...	lex-gram	colligation	clause	misselection	reformulation > those who come under the fourth power 's influence (?)	calque: podatni na działanie tej czwartej władzy
116	2027	the DIFFICULTIES the Kindergartens have TO TACKLE with (sth)...	lex-gram	colligation	clause	misselection	ing/inf > the DIFFICULTIES the Kindergartens have, TACKLING (sth)...	calque: trudności jakie mają by radzić sobie ...
117	2027	the difficulties the Kindergartens have to tackle (sic!) WITH LACKING places	lex-gram	colligation	sentence	misselection	NP structure > the difficulties Kindergartens have, tackling the LACK of places OR the difficulties with the LACK of places Kindergartens have to tackle	calque: trudności jakie przedszkola mają, by uporać się z brakującymi miejscami OR trudności z brakującymi miejscami, z którymi przedszkola muszą się zmagać
118	2005	Following this thought, it is essential to SAY ABOUT parents who...	lex-gram	colligation	phrase	misselection	object > say SOMETHING about OR it is essential to mention/consider/discuss parents who…	calque: trzeba powiedzieć o... (rodzicach)
119	2004	force children to EDUCATION	lex-gram	colligation	phrase	misselection	> force sb to DO sth (to study/to seek education?)	calque: zmusić do wykształcenia
120	2032	at the age of six there is a CHANCE to motivate students to...	lex-gram	colligation	sentence	misselection	ing/inf > a chance of motivating? (likelihood)	underdifferentiation: chance - confusion from PL "szansa" used with inf in expression of probability (e.g. miec szansę wygrać)
121	2032	families whose income does not ALLOW * to enrol their children to kindergartens	lex-gram	colligation	sentence	omission	object > families whose income does not allow them to enrol their children in kindergartens	calque: dochód nie pozwala zapisać dzieci do przedszkola
122	2029	They are given the recipes how to become...	lex-gram	colligation	sentence	omission	prep > are given instant recipes FOR how to become...	calque: mają podane recepty jak zostać…
123	2016	I do not approve * his behaviour	lex-gram	colligation	phrase	omission	prep > I do not approve OF his behaviour (covert error: approve sth - zatwierdzić)	calque: no prep in PL but inflection: nie pochwalam jego zachowania
124	2031	object * the idea of...	lex-gram	colligation	phrase	omission	prep > object TO the idea of...	calque: przeciwstawiać się (pomysłowi, aby...)
125	2014	Does winning 8 golden medals … SAY for nothing?	lex-gram	collocation	sentence	blend	> STAND for nothing/ SAY nothing/ MEAN nothing	calque: o niczym nie mówi? (two TL forms blended into one)
126	2029	...quality papers and tabloids. ... that they read only THOSE FIRST ONES	lex-gram	collocation	text	misselection	> the former? (the first ones?)	calque: "te pierwsze"
127	2002	STRESSFUL factor	lex-gram	collocation	phrase	misselection	> stress factor	calque: czynnik stresujący
128	2007	..." ,CLAIMS a well-known saying, although...	lex-gram	collocation	phrase	misselection	> as a well-known saying GOES	calque: jak mówi/głosi słynne twierdzenie
129	2013	his stupid acts may contribute to MORAL DISASTERS of young people	lex-gram	collocation	phrase	misselection	> moral decline	calque: klęska moralna
130	2033	young MINDS whose outlooks have not yet been fully shaped	lex-gram	collocation	sentence	misselection	> young people, whose outlooks have not yet been fully shaped	calque: młode umysły, których poglądy jeszcze się w pełni nie ukształtowały
131	2016	have to HAVE a toll on his personality	lex-gram	collocation	phrase	misselection	> have to take their toll on his personality	calque: muszą mieć wpływ na jego osobowość

132	2034	some children... will not be able to make CONTACT with others	lex-gram	collocation	clause	misselection	> establish a connection/develop relations ('make contact' is a collocation but with a different, more literal meaning)	calque: nawiązać kontakt (meaning "develop relations"?)
133	2025	huge amounts of money they OPERATE	lex-gram	collocation	clause	misselection	> huge amounts of money they handle	calque: ogromne ilości pieniędzy jakimi operują
134	2039	gains its peak of possibilities	lex-gram	collocation	phrase	misselection	> achieves/reaches/realizes its full potential (many more!)	calque: osiąga szczyt swoich mozliwości (a whole phrase involved here; more than underdiff)
135	2007	increase OF stress	lex-gram	collocation	phrase	misselection	prep (Polish gen?) > increased stress/increase IN stress (?)	calque: PL: genitive (wzrost/nasilenie stresu?)
136	2007	focuses on struggles which they are supposed to TAKE	lex-gram	collocation	sentence	misselection	> struggles which they are supposed to experience OR: efforts they are supposed to make	calque: podjąć trudy
137	2005	OBEDIENT to this rule I put…	lex-gram	collocation	sentence	misselection	> In line/keeping with this principle	calque: posłuszny tej zasadzie
138	2006	STAY objective	lex-gram	collocation	phrase	misselection	> remain objective	calque: pozostać obiektywnym
139	2034	the (sic!) school education should be left FOR more qualified teachers...	lex-gram	collocation	phrase	misselection	prep > school education should be left to more qualified teachers...	calque: pozostawiona bardziej wykwalifiowanym nauczycielom
140	2027	it is the citizen whom LAW is STOOD for	lex-gram	collocation	sentence	misselection	> made/created (it is the citizen that law is made/created for)	calque: prawo jest stanowione/ustalane
141	2005	THROW many arguments	lex-gram	collocation	phrase	misselection	> give/offer/present/produce arguments	calque: rzucać argumenty
142	2005	PSYCHOLOGICAL ABILITIES	lex-gram	collocation	phrase	misselection	> mental skills/abilities	calque: sprawność psychiczna
143	2026	an AWFUL amount of courage	lex-gram	collocation	phrase	misselection	> tremendous/unbelievable amount of courage	calque: strasznie dużo odwagi (?)
144	2026	Michael has UNDERGONE the consequences	lex-gram	collocation	phrase	misselection	> has suffered the consequences	calque: undergone - poniósł (konsekwencje) (similarity in morphological structure)
145	2027	not only does the government STAND law	lex-gram	collocation	clause	misselection	> make/create law	calque: ustala/ustanawia prawo
146	2034	IN young age	lex-gram	collocation	phrase	misselection	prep > at a young age	calque: w młodym wieku
147	2004	IN such a young age	lex-gram	collocation	phrase	misselection	> at such a young age	calque: w tak młodym wieku
148	2034	IN the age of six	lex-gram	collocation	phrase	misselection	prep > at the age of six	calque: w wieku...
149	2039	IN the right time	lex-gram	collocation	phrase	misselection	> AT the right time	calque: WE właściwym czasie
150	2034	see the IMMENSE distinction	lex-gram	collocation	phrase	misselection	> see the clear/sharp distinction	calque: widzą ogromą różnicę
151	2034	JOINT learning	lex-gram	collocation	phrase	misselection	> group learning	calque: wspólna nauka
152	2016	HIGH penalty	lex-gram	collocation	phrase	misselection	> heavy penalty/punishment	calque: wysoka kara
153	2007	[the government's decision] will cause a chain reaction STIMULATING mourning, disappointment or...	lex-gram	collocation	sentence	misselection	semantic prosody > bringing mourning	calque: wywołując żałobę (PL may use the word ironically)
154	2034	instead of enrolling him/her TO kindergarten	lex-gram	collocation	phrase	misselection	prep > instead of enrolling him/her in kindergarten	calque: zapisać DO przedszkola

155	2032	enrol their children TO kindergartens	lex-gram	collocation	phrase	misselection	> enrol their children in kindergartens	calque: zapisać dzieci DO przedszkola
156	2038	to enrol a child TO school	lex-gram	collocation	phrase	misselection	> to enrol a child in/at school	calque: zapisać dziecko do szkoły
157	2027	enrolling six-year-olds TO school	lex-gram	collocation	phrase	misselection	> enrolling six-year-olds in school	calque: zapisywanie sześciolatków DO szkoły
158	2016	not to USE drastic sanctions ON M.P.	lex-gram	collocation	phrase	misselection	> impose sanctions on; use sanctions against	calque: zastosować sankcje wobec
159	2034	deprive them FROM... memories	lex-gram	collocation	phrase	misselection	prep > deprive them of... memories	feature absent from L1: no prep in PL but inflection
160	2015	deprived FROM morality	lex-gram	collocation	phrase	misselection	prep > deprived of morality	feature absent from L1: no prep in PL but inflection
161	2036	IN the very beginning of...	lex-gram	collocation	phrase	misselection	> At the very beginning of...	underdifferentiation - No distinction in PL: na początku - at the beginning/in the beginning
162	2034	...which be (sic!) visible DURING his whole life	lex-gram	collocation	phrase	misselection	> throughout his life	underdifferentiation - PL: podczas > during/over/throughout/for... ("during + det + life" - OK, but not if preceded by WHOLE)
163	2015	is not going to MAKE anything stupid	lex-gram	collocation	phrase	misselection	> do	underdifferentiation - zrobić (make/do)
164	2010	more classrooms for the SOONER BEGINNERS	lex-gram	collocation	phrase	misselection	> for the early beginners	underdifferentiation: dla wcześniej zaczynających - WCZEŚNIE > early/soon ZACZYNAĆ: begin/start
165	2008	people need to discern the problem	lex-gram	collocation	phrase	misselection	> recognise the problem	underdifferentiation: dostrzec problem
166	2003	the only time people think ABOUT as unusual is…	lex-gram	collocation	phrase	misselection	prep: opinion > "think OF sth as + adj"	underdifferentiation: mysleć O czymś - PL: no distinction between opinion/activity
167	2029	who are not rarely forced	lex-gram	collocation	clause	misselection	adverb > who not infrequently are forced (OR are often/frequently forced)	underdifferentiation: nierzadko (rarely/infrequently - only the latter appears with NOT)
168	2032	may be perceived as a PEJORATIVE feature	lex-gram	collocation	phrase	misselection	> negative feature	underdifferentiation: PL - cecha negatywna negatywny: pejorative/ negative
169	2031	they are REFUSED the right to privacy	lex-gram	collocation	clause	misselection	> they are denied the right to privacy	underdifferentiation: PL - odmawiać (deny/refuse)
170	2007	they will MANAGE WITH all so challenging difficulties	lex-gram	collocation	phrase	misselection	> They will COPE with all the challenges/difficulties	underdifferentiation: poradzą sobie (z wszelkimi trudnościami) manage/cope
171	2038	a wide social discussion	lex-gram	collocation	phrase	misselection	> a broad public debate	underdifferentiation: szeroka (wide/broad) dyskusja społeczna (social/public)
172	2008	to ACQUIRE good sports RESULTS	lex-gram	collocation	phrase	misselection	> achieve good results in sport	underdifferentiation: UZYSKAĆ dobre wyniki sportowe
173	2003	a GREAT financial relief	lex-gram	collocation	phrase	misselection	> SUBSTANTIAL/SIGNIFICANT/BIG financial relief	underdifferentiation: wielka ulga finansowa
174	2008	the fact that they are creating opinion and view of many people	lex-gram	collocation	phrase	misselection	> that they are FORMING many people's opinions and views	underdifferentiation: że BUDUJĄ postawy i poglądy wielu osób (create/form/build)
175	2033	to... DRAW a helping hand	lex-gram	collocation	phrase	misselection	> extend/reach out	underdifferentiation: wyciągnąć (pomocną dłoń) (draw/extend/reach out a helping hand)
176	2033	a need of breaking this vicious circle appears to INCREASE	lex-gram	collocation	clause	misselection	> to be GROWING	underdifferentiation: wzrastać > grow/increase

177	2010	ENLARGE the stuff (sic!) number	lex-gram	collocation	phrase	misselection	> INCREAESE the number of staff	underdifferentiation: PL: zwiększyć > enlarge/increase
178	2016	setting a NEGATIVE example	lex-gram	collocation	phrase	misselection	> bad example	underdifferentiation: zły - bad/negative
179	2007	there is a tendency FOR FLEEING FOR FAME (sic!)	lex-gram	collocation / colligation	sentence	misselection	> tendency to seek / pursue fame (and fortune) (also: prep+ ingvs. inf)	calque: istnieje tendencja do pogoni za sławą
180	2007	[the government's decision] will cause a chain reaction STIMULATING MOURNING, disappointment or...	lex-gram	style	sentence	misselection	> causing/provoking/breeding bitterness/resentment	calque: wywołując żałobę (?) (figurative use in PL)
181	2005	economical PLUSSES	lex-gram	style	text	misselection	> economic advantages/benefits	underdifferentiation - "plusy" - informal!
182	2008	Phelps made a mistake but he has come to particular conclusions.	lex-gram	style	sentence	misselection	> come to some/certain conclusions	underdifferentiation - doszedł do określonych/konkretnych wniosków (covert error)
183	2036	hordes of CONTEMPORARY carefree teenagers	lex-gram	style	sentence	misselection	stylistic clash! > HORDES of present-day/today's carefree teenagers	underdifferentiation - tabuny współczesnych beztroskich nastolatków - less contrasting
184	2004	interactions with other HUMANS	lex-gram	style	text	misselection	> interactions with other people	underdifferentiation: ludzie>people/men/humans - stylistic differences/register. Fewer options in Polish
185	2034	spending * (sic!) with CONTEMPORARIES has * (sic)! significant effect on...	lex-gram	style	text	misselection	> peers (much too formal for the context)	underdifferentiation - rówieśnicy (peers/contemporaries)
186	2019	DISTINGUISHED and prominent superstar	lex-gram	style	text	misselection	> glamourous superstar ("distinguished" is used in very formal contexts, usually in reference to people of high social status in such fields as science or politics)	underdifferentiation: prawdziwa/wielka gwiazda
187	2015	if he or she WOULD BE to suffer	lex-gram	tense/aspect	clause	misselection	past condition > were/was to suffer	calque: Polish - hypothetical form required
188	2004	it would seem that this problem WAS already solved	lex-gram	tense/aspect	clause	misselection	> has already been solved (*present reference*)	feature absent from L1: perfect tense - został już rozwiązany
189	2033	a need of breaking this vicious circle appears TO INCREASE	lex-gram	tense/aspect	sentence	misselection	progressive > to BE GROWING?	feature absent from L1: zdaje się wzrastać (infinitive - no corresponding aspect distinction in PL)
190	2013	that were the swimming star not so determined, he would not MAKE such * splendid career	lex-gram	tense/aspect	text	misselection	> would not have made	underdifferentiation: no distinction between hypothetical past/present in PL: nie zrobiłby kariery
191	2008	Phelps is THE SAME PERSON as they are	lex-gram	word choice	sentence	overinclusion	confusing! (identity?) > is the same as they are OR is the same KIND OF person as they are OR he is a person JUST LIKE them	calque: jest taka samą osobą jak oni (as it is: jest TĄ samą osobą)
192	2036	as anonymous and average as A GREY MOUSE	lex-gram	word choice	sentence	misselection	> as timid as a mouse/rabbit	calque: jak szara myszka
193	2007	Not only can each students (sic!) POSSESS new knowledge and adjust...	lex-gram	word choice	text	misselection	> acquire/gain new knowlege - covert error; intended meaning: process, not state	"false friends": pozyskać/possess - POZYSKAĆ nową wiedzę (formal similarity);

194	2002	They are better … CONCEN-TRATED to start learning	lex-gram	word choice	sentence	misselection	> focused (on sth!) ("concentrated" - used as adj only attributively, with nouns like effort, attention, power)	"false friends": skoncentrowany/concentrated; surface similarity - (pp/adj not used with the infinitive for 'thinking intensely')
195	2033	the common people watch what media SERVE	lex-gram	word choice	phrase	misselection	> media present/OFFER	"false friends": to co media SERWUJĄ
196	2005	those most INTERESTED - chil-dren.	lex-gram	word choice	text	misselection	> those most concerned	calque: dla najbardziej zainteresowanych
197	2003	None of the parents wants his or her child to be an experimental rabbit	lex-gram	word choice	sentence	misselection	> guinea pig / lab rat (AmE)	calque: królik doświadczalny
198	2029	They are given the recipes how to... on the palm of their hands	lex-gram	word choice	sentence	misselection	> are given INSTANT recipes for becoming...	calque: mają podane recepty jak zostać ... jak na dłoni (English "IN the palm of one's hand" has a different meaning - "w garści")
199	2005	DEPRIVED of help from grand-parents	lex-gram	word choice	text	misselection	> with no help from...	calque: pozbawiony pomocy dziadków
200	2027	the ... PROBLEM WITH LACK-ING PLACES in kindergartens.	lex-gram	word choice	sentence	misselection	> lack of places	calque: problem z brakującymi miejscami
201	2006	the teachers have to HURRY WITH THE MATERIAL	lex-gram	word choice	text	misselection	> rush through the course (material)	calque: spieszą się z materiałem
202	2016	With regard (end of a letter)	lex-gram	word choice	text	misselection	> Regards,	calque: Z poważaniem
203	2020	the time when children experience CARELESS and safe existence	lex-gram	word choice	sentence	misselection	> carefree (sem. pros.) - covert error	coinage: "-less"/"bez-"; e.g. beznadziejny-hopeLESS - hence: beztroski > 'careless'
204	2007	deprive their children of a few years of CARELESS existance (sic!)	lex-gram	word choice	sentence	misselection	> carefree (sem. pros.) - covert error	coinage: "-less"/"bez-"; e.g. beznadziejny-hopeLESS - hence: beztroski > 'careless'
205	2003	the end of childhood, of careless-ness...	lex-gram	word choice	phrase	misselection	> lightheartedness (sem. pros.) - covert error	coinage: "-less"/"bez-"; e.g. beznadziejny-hopeLESS - hence: beztroski > 'careless' +nominal suffix -ness > carelessness
206	2039	wonderful time of total CARE-LESSNESS	lex-gram	word choice	phrase	misselection	> lightheartedness (sem. pros.) - covert error	coinage: "-less"/"bez-"; e.g. beznadziejny-hopeLESS - hence: beztroski > 'careless' +nominal suffix -ness > carelessness
207	2010	the joyful innocence and the sweet CARELESSNESS	lex-gram	word choice	phrase	misselection	> lightheartedness (sem. pros.) - covert error	coinage: "-less"/"bez-"; e.g. beznadziejny-hopeLESS - hence: beztroski > 'careless' +nominal suffix -ness > carelessness
208	2033	surpasses all that one has tried to ENLIGHTEN so far...	lex-gram	word choice	text	misselection	> to present?	coinage: PL: co próbowano tu na-świetlić >>en-lighten
209	2010	Some... ministers ... ALTERED the outfits of the students.	lex-gram	word choice	text	misselection	> dictated to students what to wear/introduced a dress code for students (covert error)	underdifferentiation - (COVERT error!)(alter vs. change clothes/outfits-przerobić ubrania!) zmienił ubiór uczniów
210	2006	provide the children with a POS-SIBILITY to socialize	lex-gram	word choice	sentence	misselection	> an opportunity to socialize	underdifferentiation - dają dzieciom mozliwość nawiązywania znajomości
211	2039	...making children attend a school earlier can harm them. Childhood is, THOUGH, not only the most beautifull...	lex-gram	word choice	text	misselection	discourse marker > after all?	underdifferentiation - Dzieciństwo jest jednak...

212	2038	older INFANTS (primary school context)	lex-gram	word choice	text	misselection	> pre-schoolers?	underdifferentiation - dziecko> infant/child
213	2005	Economical crissis (sic!)	lex-gram	word choice	phrase	misselection	> economic	underdifferentiation - kryzys ekonomiczny
214	2005	ECONOMICAL crisis	lex-gram	word choice	phrase	misselection	> economic	underdifferentiation - kryzys ekonomiczny (ekonomiczny - economic/economical)
215	2006	JOIN school classes with play	lex-gram	word choice	phrase	misselection	> mix/blend/combine	underdifferentiation - łączyć lekcje z zabawą - synonyms with one equivalent in PL
216	2016	We MAY see his photographs in almost every city around the globe.	lex-gram	word choice	sentence	misselection	> modal: can	underdifferentiation - możemy (can/may)
217	2002	creates such a POSSIBILITY for children	lex-gram	word choice	sentence	misselection	> opportunity	underdifferentiation - możliwość (possibility/opportunity/chance, etc.)
218	2039	offers pupils ... the POSSIBILITY of gaining good habits	lex-gram	word choice	sentence	misselection	> opportunity/chance to develop	underdifferentiation - możliwość (possibility/opportunity/chance, etc.)
219	2036	it is ... INDISPENSABLE to be mentioned that...	lex-gram	word choice	sentence	misselection	> necessary (different use & meaning!)	underdifferentiation - PL: konieczne - necessary/indispensable
220	2034	such a child can LEARN in a quiet and peaceful atmosphere	lex-gram	word choice	sentence	misselection	> can STUDY	underdifferentiation - PL: uczyć się>learn/study
221	2034	opportunity to LEARN with other children	lex-gram	word choice	sentence	misselection	> study	underdifferentiation - PL: uczyć się>learn/study
222	2034	do not have ELIGIBLE didactic education	lex-gram	word choice	text	misselection	> do not have sufficient educational qualifications	underdifferentiation - PL: uprawniony>eligible/qualified! (wykształcenie z uprawnieniami dydaktycznymi)
223	2027	the RAISING problem	lex-gram	word choice	phrase	misselection	> RISING	underdifferentiation - podnosić (rise/raise)
224	2005	This idea, although impossible to INTRODUCE, would be perfect because...	lex-gram	word choice	phrase	misselection	> ,although impossible to IMPLEMENT, (covert error)	underdifferentiation - pomysł ten, choć niemożliwy do zrealizowania (dosł. wprowadzenia w życie),
225	2027	As children... have the necessity to move	lex-gram	word choice	phrase	misselection	> have the need	underdifferentiation - potrzeba/konieczność
226	2020	ESPECIALLY requalified teachers	lex-gram	word choice	phrase	misselection	> specially	underdifferentiation - specjalnie - specially/especially (some overlap in Enlish usage, but not here)
227	2032	IT is due to the fact that...	lex-gram	word choice	text	misselection	> THIS (broad reference)	underdifferentiation - TO - same Polish pronoun for IT/THIS in anaphoric reference
228	2010	were pupils to LEARN longer, their results would…	lex-gram	word choice	sentence	misselection	> study	underdifferentiation - uczyć się - learn/study
229	2019	and, AT LAST, consistency which makes us reliable and respected people	lex-gram	word choice	text	misselection	> finally	underdifferentiation - w końcu (two functions impatience v. enumeration)
230	2034	do not have eligible (sic!) DIDACTIC education	lex-gram	word choice	text	misselection	> do not have sufficient educational qualifications	underdifferentiation - wykształcenie z uprawnieniami dydaktycznymi
231	2014	8 (sic!) GOLDEN medals	lex-gram	word choice	text	misselection	> gold medals	underdifferentiation - złoty > gold/golden (no such semantic distinction in PL)
232	2005	ECONOMICAL plusses	lex-gram	word choice	text	misselection	> economic advantages	underdifferentiation: ekonomiczny>economic/economical (

233	2025	news which people think to be objective SOAK with political ... lobbying	lex-gram	word choice	sentence	misselection	> is SATURATED with (?)	underdifferentiation: ociekają? są przesiąknięte? (SOAKED not used figuratively)
234	2010	(they) REFRAINED themselves only to the changes in the curriculum	lex-gram	word choice	clause	misselection	> limited themselves	underdifferentiation: ograniczyć się (refrain/limit oneself)
235	2032	proponents of this point of view	lex-gram	word choice	phrase	misselection	no evidence of such collocation > proponents of this view!	underdifferentiation: PL - punkt widzenia=opinia(?)
236	2007	the most significant part of * HUMAN's life	lex-gram	word choice	text	misselection	> a person's/an individual's life (human: used if contrasted with animals or machines)	underdifferentiation: w życiu człowieka (PL człowiek>man/person/a human being/
237	2029	But is it all so easy to STATE, that today's media...	lex-gram	word choice	sentence	misselection	underdifferentiation > But is it all so easy to SAY that today's media...	underdifferentiaition: łatwo stwierdzić (state /claim/say)
238	2029	The great majority of people STATE, that today's newspapers...	lex-gram	word choice	sentence	misselection	underdifferentiation > The great majority of people CLAIM that today's newspapers...	underdifferentiation: ogromna większośc ludzi twierdzi (state / claim)
239	2005	Nowadays citizens live in fear of losing their hard earned pensions	lex-gram	word choice	text	omission	> salaries? OR SENIOR citizens? (covert error) - (context > the former)	"false friends" (pensja/pension) OR underdifferentiation - emeryci (incomplete equivalent acquired - because two words in English?)
240	2005	SINCE I can remember it has been	lex-gram	word choice	sentence	omission	confusing: since = conjunction of reason? > EVER since I can remember	underdifferentiation: since - odkąd/ponieważ
241	2036	Aside from the model role assigned to the sport stars, it is, HOWEVER, indispensable...	synt	clause combining	sentence	overinclusion	double marking of contrast/concession > delete HOWEVER?	calque: pomimo roli wzorca, jaka przypisywana jest gwiazdom sportu, jest jednak konieczne aby...
242	2004	to remember that we can not (sic!) force children to education AND we should rather encourage them	synt	clause combining	sentence	misselection	> that we cannot force children to seek education BUT * should rather encourage them	calque: a raczej powinniśmy je zachęcać
243	2019	the way HOW top athletes set examples (sic!)	synt	clause combining	sentence	misselection	> the way (that) top athletes set an example	calque: sposób w jaki wielcy sportowcy dają przykład
244	2033	sentence length [77 words!]	synt	clause combining	sentence	misselection	sentence too long - confusing structure	Polish generally allows longer sentences
245	2033	sentence length [35 words!]	synt	clause combining	sentence	misselection	sentence too long - confusing structure	Polish generally allows longer sentences
246	2019	sentence length [53 words!]	synt	clause combining	sentence	misselection	sentence too long - confusing structure	Polish generally allows longer sentences
247	2026	LIKE Phelps did	synt	clause combining	clause	misselection	comparative/style! > as Phelps did (possible in informal style)	underdifferentiation - like/as - PL "jak"
248	2031	they storm into people's houses or hospitals and question THEM about the most tragic...	synt	clause combining	sentence	misselection	reference pron/antecedent agr - confusing > and ask questions about...	underdifferentiation - Polish - different forms of the pronouns would be involved - so no confusion
249	2006	there are also those WHICH express a strong objection	synt	clause combining	sentence	misselection	> who	underdifferentiation - są tacy, którzy wyrażają ostry sprzeciw (PL - one relative pronoun for human and non-human reference)

250	2004	the psyche of a child, WHICH could be inhibited and afraid of school	synt	clause combining	sentence	misselection	> who	underdifferentiation > dziecko, które (PL - one relative pronoun for human and non-human reference)
251	2025	examples * how news which...	synt	clause combining	sentence	omission	preposition > examples OF how...	calque: wiele przykładów jak wiadomości
252	2029	But is it all so easy to state, THAT today's media...	synt	clause combining/punct	sentence	overinclusion	comma + that > But is it all so easy to say THAT today's media...	transfer of training (in L1): PL - comma always in front of "że"
253	2029	It is claimed, THAT they read...	synt	clause combining/punct	sentence	overinclusion	comma + that > It is claimed THAT they read...	transfer of training (in L1): PL - comma always in front of "że"
254	2029	The great majority of people state, THAT today's newspapers...	synt	clause combining/punct	sentence	overinclusion	comma + that > The great majority of people claim THAT today's newspapers...	transfer of training (in L1): PL - comma always in front of "że"
255	2034	those who argue, THAT (comma)	synt	clause combining/punct	sentence	overinclusion	comma + that > those who argue THAT	transfer of training (in L1): PL - comma always in front of "że"
256	2026	We all make mistakes, after all, nobody is perfect.	synt	clause combining/punct	sentence	misselection	comma splice >... ; after all... OR a new sentence: ... . After all...	transfer of learning in L1: Polish usage rules are much less strict about comma splice, and commonly ignored in schools
257	2006	all costs of a current elementary school THAT includes only six grades	synt	clause combining/punct	sentence	misselection	comma! + which > all costs of a current elementary school, which includes only six grades	underdifferentiation - dzisiejszej szkoły podstawowej, która... (PL - one relative pronoun, no distinction between defining and non-defining relative clauses)
258	2026	an awful amount of courage, THAT most athletes do not risk showing.	synt	clause combining/punct	sentence	misselection	non-def rel. cl. > a tremendous/massive amount of courage, which most athletes do not risk showing.	underdifferentiation - PL - no distinction
259	2016	He should be well aware of the consequences ... , however, we should take into consideration that he is only human.	synt	clause combining/punct	sentence	misselection	comma splice (contrast/concession) > although OR but OR start a new sentence with 'However'	underdifferentiation: PL 'jednakże' may be used as an adverb or as a subordinator
260	2005	Since I can remember it has been at the age of seven, however, now ... many controversies arose.	synt	clause combining/punct	sentence	misselection	comma splice (contrast/concession) > although OR but OR start a new sentence with 'However'	underdifferentiation: PL 'jednakże' may be used as an adverb or as a subordinator. If adverb, then a new sentence must begin with 'however'.
261	2002	The decision was motivated by a number of economic and social issues, however, not all of them were convincing.	synt	clause combining/punct	sentence	misselection	comma splice (contrast/concession) > although OR but OR start a new sentence with 'However'	underdifferentiation: PL 'jednakże' may be used as an adverb or as a subordinator
262	2032	Many of them HAVE NOT enough facilities	synt	clause structure	sentence	misordering	negative > DO NOT HAVE enough facilities	calque: ma nie dość sprzętu
263	2027	such young learners CAN BE NOT prepeared to...	synt	clause structure	sentence	misselection	negative > such young learners may not be prepared to... /may be unprepeared to.../	calque: mogą być nieprzygotowani
264	2033	this is what they would surely long for if * found themselves drowning	synt	clause structure	sentence	omission	sub subject > this is what they would surely long for if THEY found themselves drowning	calque: gdyby sami tonęli

265	2015	how chalenging (sic!) * is to be a professional journalist	synt	clause structure	sentence	omission	sub subject > how challenging it is to be a professional journalist	calque: jak trudno jest być zawodowym dziennikarzem
266	2003	Besides economic advantages social * can be distinguished	synt	clause structure	sentence	omission	> Besides economic advantages, social ONES can be distinguished - ellipsis instead of substitution	calque: oprócz zysków gospodarczych można też zauważyć społeczne - ellipsis much more common in Polish
267	2002	sending their children * one year earlier	synt	clause structure	clause	omission	complex transitive - missing adverb > sending their children TO SCHOOL one year earlier	calque: posyłanie dzieci wcześniej
268	2003	the child is better prepared … because * has started earlier	synt	clause structure	sentence	omission	sub subject > the child is better prepared … because HE/SHE has started earlier	calque: silent subject - "ponieważ * zaczęło wcześniej"
269	2016	I would prefare (sic!) * if icons were true to themselves...	synt	clause structure	sentence	omission	missing object > I would prefer IT if icons were... OR better: I would rather icons were...	calque: wolałbym, gdyby ikony były
270	2005	because discouraged at the very beginning * will not learn	synt	clause structure	sentence	omission	sub subject > because discouraged at the very beginning he will not learn	calque: zniechęcony na początku nie będzie się uczył...
271	2033	separating what is true from FALSE	synt	parallel	sentence	misselection	> separating what is true from what is false	calque: od fałszu?
272	2019	should be aware OF his enormous influence... and THAT...	synt	parallel	sentence	misselection	> English more strict re parallel structure > aware that he has.. and that...	calque: Powinien być świadomy swego ogromnego wpływu i że...
273	2016	I would prefare (sic!) * if icons were true to themselves than fake	synt	parallel	sentence	omission	> I would prefer it if icons were true to themselves RATHER than fake. OR : I would rather icons were true ..., not fake.	calque: wolałbym, gdyby ikony były wierne sobie, niż fałszywe.
274	2034	EDUCATING BY PARENTS was very popular...	synt	voice	clause	blend	active/passive confusion (by - used in passive) > Parents educating children were very... (BETTER: home schooling!)	calque: kształcenie przez rodziców - PL OK
275	2019	remember that their job does not limit to their sports achievements	synt	voice	sentence	misselection	passive/active > is not limited to their sports achievements	calque: ich praca nie ogranicza się do...
276	2025	they raise questions of invading one's privacy	synt	voice	clause	misselection	agent NOT subject > invasion of privacy?/passive ger? N+ger? (people's privacy being invaded/journalists invading...)	calque: podnoszą kwestie naruszania czyjejś prywatności
277	2025	news which people think to be objective SOAK with political ... lobby	synt	voice	sentence	misselection	> IS SATURATED with	calque: wiadomości ociekają?
278	2004	Not only it would have a negative (sic!) effect ... but also it would...	synt	wo	sentence	misordering	inv > Not only would it have...	calque: (no similar emphatic mechanism in PL) Nie tylko miałby...
279	2010	it will be as well enriching	synt	wo	sentence	misordering	> it will be enriching as well	calque: będzie to też wzbogacające - less rigid WO rules (position of adverb)
280	2003	keep all the time up to date	synt	wo	sentence	misordering	time adv separates V from Complement > keep up to date all the time	calque: być cały czas na bieżąco

281	2007	..." , claims a well-known saying, although...	synt	wo	sentence	misordering	> as a well-known saying goes	calque: głosi słynne twierdzenie
282	2004	are more creative, WHICH TRAIT helps them...	synt	wo	sentence	misordering	possible but very marked - not recommended unless the remaining text is flawless > are more creative, a trait which helps them...	calque: która to cecha...
283	2003	taxes shouldn't be here forgotten	synt	wo	sentence	misordering	adv position > taxes should not be forgotten here	calque: nie należy tu zapominać o podatkach
284	2013	that M.P. has a special and vital role to play in society as well as other high profile athletes.	synt	wo	sentence	misordering	> that M.P., as well as…, has a special…	calque: PL - more flexible WO rules (cases!)
285	2031	the most problematic media's target	synt	wo	phrase	misordering	superlative + possesive > media's most problematic target?)	calque: PL - superlative - phrase initial
286	2005	the earlier they will (HOPEFULLY) finish it	synt	wo	sentence	misordering	position of a viewpoint adverb >...the earlier they will finsh it, hopefully/let us hope.	calque: tym wczesniej, miejmy nadzeję, ją skończą
287	2003	Mentioned above aspects are …	synt	wo	sentence	misordering	noun pre-/post-modification > The aspects mentioned above are...	calque: Wyżej wspomniane aspekty... (PL: premodification of NP is far less structurally constrained)
288	2003	not only the child is better prepared ... but also brings more profits to its country.	synt	wo	sentence	misordering	inv > not only is the child better prepared	feature absent from L1: (no similar emphatic mechanism in PL) Nie dość to, że dziecko jest lepiej przygotowane...
289	2004	Not only children develop their social skills but also they are creative and…	synt	wo	sentence	misordering	inv > Not only do children develop their social skills but also… (covert error!)	feature absent from L1: (no similar emphatic mechanism in PL) Nie tylko dzieci rozwijają... ale i ...
290	2008	a WORLD-WIDE KNOWN superstar	synt	wo	sentence	misordering	noun pre-/post-modification > a superstar who is known world-wide OR BETTER: an INTERNATIONAL superstar	calque: znana na całym świecie gwiazda
291	2004	Not only they recive (sic!) better grades but also are more creative	synt	wo	sentence	misordering	inv > Not only do they receive better grades...	feature absent from L1: (no similar emphatic mechanism in PL) nie tylko otrzymują lepsze oceny…
292	2003	regarding how children should … and when should start their education	synt	wo	sentence	omission/misordering	imbedded quest. OR sub subject omission > when their education should start OR when they should start...	calque: kiedy powinna zacząć się ich edukacja

Appendix B: Blog corpus search results

‘-ly’ search - 3589 hits, 80 errors (here – errors only)

1. is reading it. Anyway, I'm still alive (suprisingly to me :P), earning my own money :). My life is st
2. lem with buying the rest staff she needed (thankully). I spend a very good time there although I don't
3. nt to join me? You are most welcome. (Unfortunatelly, there is only one direct rail connection between
4. t know if you are aware of the fact that I acctually come from the coolest city in the whole Poland...
5. els is seemingly real. Some critics have accuratelly labelled his style as a magic realism. But his wo
6. ill: There are currently 3 film genres alternatingly present in our country. These are: 1. Romantic Co
7. etty bur have a lot of self-confidence and aparently a great talent. Luckily one day I discovered the
8. ears and Paris Hilton survive in this world. Basicly they are blond, not so pretty bur have a lot of s
9. gine my life without a computer because I basicially do everything with the help of my computer. Post
10. t of later-to-be film stars, with Ethan Hawke bearly out of highschool. Touching and inspiring. BTW, I
11. ch Zu resembles me :) Hope she won't be so chronicly afraid of camera as I used to be ;) My hairy-fa
12. late). Anyway, by this time I had changed competely. Well, the school changed me. I became an extreme
13. weather was really wonderful. The sky was complately cloudless and the sun was shining all the time. W
14. nd every SINGLE word - my attitude changed completly. I only wish I had more time for reading...... S
15. nt cultural events due to which you can be constanly aware of the culture around you. I am sure that c
16. ribute to my own development?"..What I'm constatntly looking for in my life are challenging people, ch
17. ice patrols were on the beat watching continuouously for any evidence of thoughtcrime. All this was to
18. any serial murderers answered this question corectly. If your answer was wrong - better for you. If an
19. and legends as well as foreign languages. He crafily interlaced this knowlegde with the stories his ma
20. body" who will probably read this post is defenitely going to kill me;) after it has turned up that I
21. harming but maybe a little bit toller and defenitely more chatty ;) The first time with IT For some p
22. Szczecinek about three years ago I would defenitely state that it's all rubbish. Whereas now, I feel
23. d girl (but still joyful and sociable). I definately suffered (and still suffer, both as a student and
24. art. If you like The Green Mile, you will definately appreciate this one. And Bjork's performance was
25. e this year, that's a fact. But it's been definately the most edifying period in my life. Great experi
26. Well, I don't remember it very well as it definately wasn't a very significant moment in my life:) How
27. bout my future especially in five years I definately would like to finish the Teacher's Training Colle
28. with najŸlejszym z³em ze z³ych z³ow which definately is KJO :P Without You I would be dead by now or a
29. ions... :D Five years later... Yes! I'm definately happy about my life! I have more than I ever want
30. ppear it is like "uuuuuuuuuu" :D This is definately my beloved city with unique spirit so come and ch
31. er table ;) I've been so full that I will definetely not manage to eat anthing else ;) It seems a good
32. have been there as it was an experience definetetly worth remembering. I epecially admired Dr Grochol
33. ual clients to help them with training. I definietly enjoy my work with animals and I guess I'm lucky
34. go on with the description. It's really difficuly to verbalise the notion of my hometown, the one t
35. xperience definetetly worth remembering. I epecially admired Dr Grocholewska's speech- it was very war

bloglines Welcome to Paradise :) I'm an exteremely beautiful princess with lots of creativity lookin
 mine? I struggled with the same dilemma, and finaly took the advice of our teacher, which was to writ
..We actually started doing something and we finnaly GOT IT! :D Dagna got it as well and I was so pr
recent suggestion (drugs taking ;) but, fortunatelly I belive, I came across a website devoted to coff
 it was extremely depressing. Silly me. Fortunatelly, God gave us Sundays when we can kneel before the
rs before and I liked it very much) :). Fortunatelly, I was wrong. So I was sitting in a room slashing
 10 times and I still like watching it. Fortunatelly I've go the director's cut version that is someth
'm far more messed up than I thought!!! Fortunatelly I don't have huge problems with it as I just have
s with the software and the whole stuff:(Fortuntely, my disc's capacity was 8 GB- that was something
ts 'cause nobody reads my blog anymore. Fortyunately I was wrong. Been asking for a full time job paid
ur phrases in Polish and English,mostly gramatically incorrect,and for fighting with najŸlejszym z³em
et rid of people who make me sick! I have some holly time to focus on my emotions, nerves, thoughts...
my heart for this place... and the painting of Holly Mary. There is a special moment when your group c
 Sunday. And I don't mean the teachings of the Holly Bible, I mean the opinion of the priests. It real
. (Polish) I wrote this poem when I was holplessly in love and I was pissed off :). (Polish) I wro
otographs. Cannot believe it! Oh my word! I honstly cannot believe in the news about John's death! He
 and (yeah!!!) as much water as I want to:) Hopefuly, one day (=soon) I'll be satisfied with my reflec
 'aktualny'; some of them can be used interchangebly while others only in ceratian contexts (depending
doing this. Sometimes I'm exhausted not only mentaly but olso physically. What can you tell when they
news for those of you who, like me, not neccessarily are looking forward to Easter . Somebody has eate
.g. Czy znasz jego aktualny adres? (not neccessarily changed) - Do you know his present address? 4. n
te music anyway, I like one tune, but not necesarily the musician as such, or the kind of music. I als
eople who have read all the books(and not necessarly understand anything) . You don't need to say than
come jealous, mean , stupid, crazy, and outragiously out of this Earth. And all the time you can feel
e computer at my home is on, this means that probaly my brother has turned it on. I seldom do it unles
mber the situation when the photo was taken, I realy like it. I seem to be so carefree, happy and safe
 anticipation of sleepless nights and, respectivelly, hurting thumbs and buttons. What has change in
and my sister to go to bed and the parents secretely packed the presents. On the Christmas Eve, my bot
m tired of hiding it form people I love but I sipmly can't tell them. The thing is that it takes a lot
ed the best and behaved the best; I was giving smily faces, suns or flowers, and on Easter lessons - c
act that the original was good enough to succesfully keep me and my dutch pal (hi Sjord!) from studyin
:)Then I will go to school and teach EFL succesfully by applying the knowlegde I received in TTT:))) A
tys. It was such an event in my 'district':) Sudenly mu cousins started visiting me and we've played g
ut by the content of [his] character". I am truely proud of the American people that they have chose
epublika.pl Has the nightmare vision of ulitimately totalitarian rule as depicted in “1984” by George

71. ger sister had been there as well but unfortunatelly it was impossible;(I've met many interesting peo
72. now somebody and than judge them. But unfortunatelly it's the other way around. Sometimes I think abou
73. all there. And Emma Frost....wow, but unfortunatelly she's not in a movie :(. I hated "The Punisher" o
74. ewrite my holiday memoirs. In Polish, unfortunatelly. It's been like three months now, and I haven't f
75. tale.. ;\ Let's get back to reality, unfortunatelly;(It was such a nice day today. The sun was shin
76. ome particular line has been crossed. Unfortunatelly. I also admire few teachers from my past schoolin
77. a lot of the material during lessons. Unfortunatelly, I was forced to make a test;(Testing is what I
78. ot what I had wanted for my birthday. Unfortunatelly, I don't remember my whish which I made while blo
79. , for example, the high unemployment. Unfortunatelly...;(Drezdenko's license plates are FSD. Do you k
80. e no matter how the whole 'event' ends. It is vitaly important in order to reach, at least the edge, o

exten(t|d) 9 hits, 7 obligatory contexts, 4 errors (undrelined)

1. single day of it! Thank you... I would like to extend my enormous thanks to my real friends by mean
2. gh the link added above;)! I do appreciate it and extend my thanks! ;-D Well done Marta! You're great!
3. lower level but free of stress at least to lesser extend. Now I have some time to think about. Send me
4. derstood they are by editing their photos to such extend their friends comment them with "wow, you loo
5. yed the first part of the film from 1988, to such extend that Aslan is still my favourite character;)
6. ry of the music that changed my life and, to some extend, shaped the person I am now. It’s all begun
7. eady frustrated with life. I believe that to some extent it is because of such situations as the one d
8. ways wanted to be an English teacher. And to some extent this does not change but I also think about d
9. ame architectonic style, same urban idea, to some extent same spirit. But still, I would like you to k

SINGULAR AND IRREGULAR PLURAL NOUNS

1. characters, Hanna, Rebeka and Leila (who’s American’s wife and is traveling with Hanna to get the money
2. I remember that the picture was taken at my auntie's Beata place. I have really good memories connecte
3. I would probably end up like the guy from Butterfly's effect, who instead of fixing, made things even w
4. startled facial expression, I wasn't used to camera's flashes ;) How proud I was of the Cracovia disgui
5. ut I can only dimly recall what each of the Carroll's novels are about. All of them are so unpredictabl
6. have one at home. When I first touched a computer's keyboard I was six or seven years old. It was my
7. My first computer's experience:) Oh my God, it was so long ago:) I re
8. It was the first time I had seen so many stage dive's, sing alongs and head walkings on one show. When
9. and lived happily after. In XVII century the Gąska’s House was added where in 1656 Karol Gustaw lived.
10. id something like: 'Girls! Please go to the grandma's room and look for the Santa'. Our elder cousin An
11. le! I perfectly remember when I went to hairdresser’s, with my Mum, and asked for a fringe. Having a fa
12.), banks, restaurants, housing estates, hairdresser’s and solariums. I like its new image, as it is mor
13. sting. I really enjoyed it (the reading, not the HP's story) and it helped me much in terms of reading
14. the way she did it, another time I admired my math’s teacher’s patience and the way he taught us, I ad
15. ours' worth of shopping, a new dress and a McDonald's vanilla milkshake to make me NOT miss certain peo
16. dreams will come true... :) A New Era? The Obama’s oath was undoubtedly historical moment. The USA h
17. idency ends. I'm rather pessimistic about the Obama's presidency. I think Americans are too excited abo
18. . I think Americans are too excited about the Obama's presidency. Nothing is going to change, especiall
19. of my favorites because it reminds me of the Orwell's blurry and terrifying Ninety Eighty Four story. H
20. N-G !!!:):) But most of all I liked the Shakespeare's teatre "The Globe"!!! I watched a play called "Un
21. Orzechowski’s implicitly calling the Polish Teacher’s Union (ZNP) communists, and the Minister of Educa
22. years I definately would like to finish the Teacher's Training College :P Just kidding, I hope it will
23. ool kids of death - and many many many more. thank's for attention. should u have any questions you kn
24. e trip was a nightmare - 3,5h standing in the train's corridor, than standing 1,5 h in a queue to check
25. ...Their arms, legs and heads I made a ton of yo yo's--No request for them, They want computers and rob

PLURAL NOUNS

1. he whole problem appeared when we went to “pancakes’ corner” - a place where you can eat noodles and pa
2. introdue you to... there are two wonderful artists' worth recommending to you all. It is Elaine Paige
3. , I go to my second work- I'm a groomer at the dogs' centre, there is also a vet, a shop with accessoir
4. r everybody. On my webpage you can find HotPotatoes' exercises for practice of our favourite phrasal ve
5. me, but he was simply astonished by a view from its' window- I wonder...). No matter what kind of a per
6. somewhere far from the noise of a big city and its' grey streets. Dreaming is the only thing I can aff
7. erful and can be recommended not only to war movies' fans. five years from now I won't write anything
8. ey the warmest feelings. I teach them to say others' eyes I like them. Or respect them. Or love them.

‘-ic’ adjectives (errors only)

ów's Nowa Huta, all of them share same architectonic style, same urban idea, to some extent same spiri
nt. how can we expect our society to present a civic maturity and responsibility, if such things are n
‘full equality’ by Black-Americans related to civic rights is not tantamount to admitting that racial
which I love, but my first choice was etnolinguistic. When it turned out that I’ll study it extramural
with over 12 tons of equipment. It's a huge logistic task, yet they do it perfectly. If not, guess who
was still great!!! I could feel that, kind of, magic atmosphere... I felt as if I was in Shakespeare's
got fantastic toys. I remember the atmosphere, magic and full of happiness. It was really nice. Now my
heck it out. My fairyland My wonderland is a magic place where all dreams come true... Once you visi
. But you will be probably charmed by the most magic place.... The Old Market.... ;) My wonderland is
till believing in Santa. I really enjoyed that magic time when I didn't have to go shopping, clean, wa
iastowski Trail, not far from many popular touristic places. It is easy to get from here to many very
easy to get from here to many very popular touristic spots such as: Kruszwica with Misia Tower, Gniezn

‘-ics’ – nouns (errors only)

, check out this video. video Daft Punk- cllasics of electronic music, "Da funk": one of the best v
 is a Pulitzer Prize winner for his incredible comics entitled 'MAUS'. The comics tells the story of Sp
for his incredible comics entitled 'MAUS'. The comics tells the story of Spiegalman's father, W³adek Sp
adek Spiegelman, former Aushwitz prisoner. The comics is a masterpiece as it presents Holocaust in the
r this pesimistic post but sometimes even optimistics have to rest and feel sad :). Medusas Again Wel
et’s face the truth – I wasn’t admitted to pedagogics. All in all, it turned out to be an ideal place f
lan to continue my education in elementary pedagogics, so my teaching practice haven’t made me change m
glish and two years later I obtained MA in Pedagogics. As a result, I am qualified to teach children at
en I go to meet God in the Church and I meet politics and political lobby instead... Why does it have t

LEVEL

1. I’d like to stay in touch with English at certain level and be more motivated to work. You know, college
2. essons I'm going to conduct. (The class is on FCE level.) But all she told me was: "Ok, so you have to do
3. sadness and melancholy. Today's hectic and high level of stress is having an extreme effect on people.
4. ust the job is killing me. Maybe if it was higher level, like LO or something. It's not that I don't like
5. GET THERE) or wsjo, which is of course on lower level but free of stress at least to lesser extend. Now
6. ... be a teacher. and if, then never for primary level... and I was supposed to teach English in German.

MAJORITY

1. future life. To put it in a nutshell, as probably majority of you (who read my blog), I would like to per
2. naged to return in full glory. First they secured majority in the US Senate and now their candidate - Bar
3. r sister... Where am I?:) With my cousins... Now, majority of us is twenty-something. The girl holding th

JOB

1. and is generally happy with her life. Of course, job which only gives you money is not most important
2. enelope Cruz looks thrilling there and does great job but I wouldn't call it a comedy.I like Woody Alle
3. rom abroad new companies have opened and there is job for almost everyone. It is possible to study at t
4. t waking up too early is. Generally, I guess that job isn`t that important for me but rather family. I
5. out! And it's all thanks to discipline... Summer job:-) Finally I got the contract:-) surprise, surpri

TEACHER

1. After some time, however, observing my Betreuer (teacher taking care of me) I became more confident. Man
2. s totally unfair. I believe all students look for teacher's attention and appreciation. Soemtimes they ju
3. anting to offend anyone, having a sense of humour teacher. I was priviliged to be his student. He always
4. and, to put it bluntly, incompetent kindergarten teacher. I stood up for my rights. It’s so funny how pr
5. would like to express my doubts about the role of teacher in class presented in the article. The author w
6. be "nasza pani Justynka" and receive flowers on teacher's day. I miss you guys! GOOD LUCK fellow teache
7. -energetic, talkative, loving children pre-school teacher who strokes your nose to wake you up - I really
8. ned to me and said, 'You are New York City school teacher. Be honest. What do you make?'. You wanna know
9. nicer... ...with a pint of bitter:D sCOOLtrip Teacher's job.... I have just returned from travel agen
10. stood on the corner of the street heard shouting teacher) caught me. I will never forget his words “ Pan
11. or a CHANGE;) Teaching practice We all know that teacher's job is not very easy. Teachers must meet all

SUCH (a) + NOUN

1. eems that one cannot find anything interesting in such area, there are some aspects worth mentioning. Fi
2. of the candidate as well as his readiness to face such challenge was the least important factor for the
3. appened that I spent almost all my life living in such district, called Rataje. It is a huge complex of
4. enjoyed the first part of the film from 1988, to such extend that Aslan is still my favourite character
5. misunderstood they are by editing their photos to such extend their friends comment them with "wow, you
6. front of all guests...I assume all the girls had such feeling...:) Fortunately, everything was OK! I wi
7. f you wish. I do not approve public flattering in such form, I do not approve attacking with words as it
8. ile we were fighting - then all of them envied me such great weapon! I also remember the year 1990.... A
9. I quarreled with my boyfriend. He couldn’t accept such idea of me having joy without him. He had many st
10. s" or "hell yeah!". Why did I even bother to make such survey? Well, to make a long story short, it's a
11. ut using poetry to develop productive skills. Why such topic? Search me! O.K., I take back my words… Ye

[have] difficulty/difficulties

1. .They have such unique voices and honestly I have difficulty in understanding the message of their lyrics:P
2. hing you like to my humble person. ;> I had some difficulty selecting my favorite movies, but in the end I
3. t in Polish and English now I'm trying (with some difficulty to be honest :) :) to read it also in spanish
4. rnate tasks which cover skills that students have difficulty with. The author gives an example of working i
5.) Oh, and one more thing... Whenever you get into difficulty, remember that "Every cloud has a silver linin
6. 2. current (girl-friend, price, rate, problems, difficulty, worry, task, job, dispute, negotiations, cont
7. role she is assigned, she fulfills is it without difficulty. Whether it is a comedy like "Mamma Mia", or a
8. nd over my knowledge which was gained in TTC with difficulty... I think everyday: 'Was it worth to slog and
9. accommodation problems - my body always had SOME difficulties adjusting back to stinky city air and a tota
10. reed to do it(Thank you so much:) But despite the difficulties I had the lesson turned out to be quite succ
11. awful, I am weak and there are days where I have difficulties just getting out of bed in the morning, not
12. e the first book at the age of 3 when I still had difficulties placing letters on paper. There are crying a
13. I must admitt that I actually did encounter some difficulties so I guess that it's a point for you Mrs. Le
14. s me of my school years, without the problems and difficulties that a mature life brings. Enjoy! Muto Th
15. that job doesn't pay. 3. present (address, job, difficulties) E.g. Czy znasz jego aktualny adres? (not ne
16. d down but it was only a short break. Despite the difficulties, I managed somehow to finish the lesson. How
17. chistic but I feel most alive when I have to face difficulties, when something's happening, when life doesn
18. ough enthusiasm and passion in spite of different difficulties. Now I'm only counting the days and I can ha

APPROVE

1. on't. Feel free to comment, if you wish. I do not approve public flattering in such form, I do not approve
2. approve public flattering in such form, I do not approve attacking with words as it has been done there,
3. ng with words as it has been done there, I do not approve making people feeling sorry for what is not thei
4. Especially in case of . But I don't have to approve everything, and I do not have to be pleased with
5. ss. i understand to some degree, although i don't approve of, the reasons standing behind the ban on "prep
6. think that he was an individual. I usually don't approve of delving into other people's lives and knowing
7. ents, and oh, how wonderful. And although I don't approve of Google's self-censoring for China, I think it
8. thought was the love of my life. My family never approved of him and when I left our contact broke comple
9. lbums - to me that's enough. ;) Besides, I really approve of the fact that he does what he feels like doin

ALLOW (errors underlined)

1. in a kindergarden, where other children where not allowed to go. I knew about a hole in the fence, so I a
2. sentation for all of the students, but time never allowed.. First of all, I was the lucky enough to take
3. being in good terms with yourself 2. Sincerity - allows to live without unclear situations which can be
4. lection; only the librarian and his assistant are allowed into the labyrinthine stacks. The ostensible re
5. as just called me and said that his parents would allow him to take their car so that we can go to Bielsk
6. e likely I am annoyed by it. I just felt the song allows to understand me a little better. If it does, mi
7. aybe 5 poems there, well, 7 at the most, if we're allowed to treat song lyrics as poetry:). I think I onc
8. another trip), so I'm not sure whether I will be allowed to get aboard the plane. Anyway, I finally mana
9. responsibility of looking after them and I can't allow myself for any hostile feelings. I need to be hap
10. t makes me want to go there as often as I can and allows me to brave the streets crowded with tourists (i
11. really life???) is a b(it difficult). At least it allows me to enjoy the soft and warm caresses of sun ra
12. the digital cameras have function "macro", which allows for such close-ups. Obviously a special lenses c
13. especially for DSLRs) or extension tubes that can allow for taking pictures from even a closer distances
14. d some of them even the same size), large buffers allowing for extremely fast shooting with the rate of 3
15. at some headmasters of high schools don't want to allow students to recors their learners as they are afr
16. xcessive demand for obedience poses. Independence allows for wider perspective - constant evaluation and
17. n. What I see as a gross dereliction, however, is allowing Poland to stand a trial in the European Court
18. e. The picture reminds me of the time, when I was allowed to sit on my Dad's shoulders and be "the talles
19. teacher is really patient and understanding. She allows us to browse the net in search for pictures of "
20. rce for teaching children is www.esl-kids.com. It allowes a teacher to create his/her own activities usin
21. . 1988? or 1989 i don't remember when was i first allowed to play games on our wonderful, beloved zx spec
22. t do. My feelings will not be repressed. You must allow me to tell you how ardently I admire and love you
23. things around me... Attention! Error correction allowed and actually very welcome. Please, let me know
24. Poland. For the first time in my short life i was allowed to cast a vote. Feeling the immense responsibil
25. worth it. Before I get to the proper show, I'll allow myself to bitch a little about the bad aspects of
26. lf. Not in my experience at least. And apart from allowing easier driver installation, easier internet ac
27. computer. It started out with some software that allowed to mix together sound samples. The catch was, t
28. onds were hard to come by. Later, better hardware allowed getting Fallout to run faster (a thorough essay
29. could "remember" up to 10 phone numbers!), I was allowed to do the first phone call - it was to my grand
30. ton please, press the star all interpretations allowed ;) and: some photos of By-dgoszcz By Zofia ;)
31. sh better relationship with Poland and we will be allowed to go to the USA without visas. I wish you good
32. RILYN MANSON band. Unfortunately, youtube doesn`t allow me to download any videos of that band, so a pict
33. king gingerbread with my Mum. I loved it that she allowed me to make my own bell- or Santa-shaped ginger

CONTACT (noun) – errors underlined

1. family never approved of him and when I left our contact broke completely. Few months passed I got a job
2. change, to guess what they are saying. I like eye contact when I feel safe and I avoid it when I feel insec
3. lame. As far back as I recall, I think my first contact with a computer was an old who-knows-what that st
4. is something what I also wish for: to still have contact with all my friends and with you guys:) Just One
5. ife, but I am not one of them. The first seriouse contact with IT started when I was in high shool and even
6. lities connected with the university I don't have contact with my instruments for days, which saddens me ve
7. and more. He is a real performer. He has a great contact with the audience, even sitting in front of the T
8. My first contact with the computer... My first contact with the computer was when I was about ten. A fri
9. or watch bbc Tv online. Have fun! :-) My first contact with the computer... My first contact with the c
10. n fact, was not mine. Certainly that was my first contact, my first impression, I was stunned watching my
11. ayout, can be put under certain labels, and [t]he contacts can be put into groups, they offer the pot
12. est e-mail ever. Everything from the inbox to the contacts has its clear layout, can be put under certain
13. addition to this, my child is keen on making new contacts with children at the same age :-D 11.30 a.m. My

IN + TIME (span:2) 44 results, errors only

1. h yeah!! :) I did it! I managed to write my B.A. IN TIME :) despite all the obstacles (my leg of cour (ON)
2. able to lead the world's most influential country IN TIME of a huge economic crisis ? I regret to say (AT)
3. he seaside. As you can see I was quite chubby but IN TIME this chubbiness has faded away... if someb (WITH)
4. "health problems". Quite unexpectedly, quite not IN TIME. Let's forget that this decision followe (TIMELY)
5. was born in August so it must have been somewhere IN THAT TIME. We call the picture "Młoda the obs (AROUND)

CARELESS (errors underlines)

1. s full of exam burden but the heart is joyful and careless :) Oh life! Best wishes to everyone! TTC pics
2. cause it had a lot of pretty pictures... I am SO careless, that's ME. I am sorry, I know this blog was sup
3. who has a broken glass, or a glass that has been carelessly knocked over (usually by one of the people cal
4. ally sentimental. I miss my childhood, mostly the carelessness and unlimited freedom. I detest Sundays… I
5. that I look at them I realize how much I miss the carelessness of the young age... The only problem was to

POSSIBILITY (errors)

1. a Polish film "Zmruż Oczy" . However, there is a possibility that I might have changed it and adopted for
2. autofocusing, easy access to all it's functions, possibility of changing lenses, better image quality due
3. ell...aspirations...ambitions...desires...so many possibiliteies...and many limitations...many things that
4. st person. I was quite nervous.... Now I have no possibility to feel the Christmas atmosphere.....I miss (I'm unable to...)
5. e my life without the computer. It gives a lot of possibilities especially when you have an access to the
6. For young people Mogilno also offers a variety of possibilities to spend leasure time. Even though Mogilno (opportunities)
7. d helping her with baking as there were plenty of possibilities to sneak something. Then I remember that I (opportunities)
8. teacher in future. But if I don't have any other possibilities (not connected with tourism) I will take t
9. ave any plans to be a teacher. I was open to that possibility, though not sure. There was a moment when I
10. But the best thing of all is the Internet and the possibility to keep in touch with my foreign friend?. (ability/that I can...)
11. eally, comes easier than writing a diary. And the possibility of putting here short films, pictures, music (ability to put)
12. ormance in my earlier post. Can he win? Well, the possibility is always there, but I already has my favori
13. Maybe it was just a matter of luck, but I had the possibility to work with a group of cheerful, involved (an opportunity)

SOME + SINGULAR COUNTABLE NOUN (errors underlined)

1. n the wings of wind Recently I was going through some blog of a person I used to know. I found this sho
2. Garland. Just before I left the library I grabbed some book by William Wharton. My English must have bee
3. atory (or at least lucky) for churchgoers to wear some bright new piece of clothing - at least an Easter
4. eam!). All the photos were taken in Poznañ during some championship (October 19). First words I could
5. r years of watching our national team qualify for some championship and f@%# it up in the first round, d
6. of music I wanted to share- Enya, Edith Piaf and some classical one. I've combined it with my another p
7. was only a link in my post ;-) Now it's time for some comment. Living on Earth is a nice web page (It
8. would do without them. Still, it worries me that some day our paths will diverge and everything will en
9. (at least for some of you). Who knows..... maybe some day, Someone up there (or wherever else) will rep
10. sue from the educational process. i understand to some degree, although i don't approve of, the reasons
11. that I was - maybe I've been - lucky, at least to some degree. This luck largely boils down to the fact
12. get my hopes up or look forward to something, to some dream and then it bursts like a baloon. But Ok I
13. ophisticated, high-end DSLR but still, If you put some effort, you can still take quite nice photos with
14. ould feel its MESSAGE as well. Now, let's provide some explanation concerning all these unusual effects.
15. hat it must have been taken in kindergarten after some fancy-dress ball. I don't know why, but tha photo
16. I want a dog or a cat. I have a huge need to hug some furry animal. Yesterday, my neighbour told me he
17. life! Best wishes to everyone! TTC pics part 3 Some guy took my scarf and ran away! If you happen to
18. rridor or carrying this huge, heavy TV to show us some interesting report and often having trouble to fi
19. NNING. That's ok, grinning is good, that means on SOME level, you're laughing. :) Alrighty, I guess thi
20. hem on the windowsill,and then they were taken in some magical way during the night :D Now something ab
21. cheer meself up. ;) I was actually offered one at some men's suit shop in Auchan Swadzim, but I am reluc
22. all over the US and touched Europe. People expect some miracle, and I hope he will not give up his plans
23. ar (October2005) to Turkey, I knew that there was some mutual animosity between the members of my own gr
24. ope for the best and I hope the ceremony breathed some new life into those most 'resistant'... Great tha
25. ike (I hope it happens soon). Today I'll chcecked some new route somewhere between Miłostowo and Antonin
26. I am in some kind of poetic mood I wanted to put some nice poem I found, looking for some works of Mr.
27. ll be close to nothing... If you set sharpness on some not important object that is 20 cm behind (or in
28. them jobs from the exact opposite ones? So after some odd order about furnaces and whichever god only k
29. ept for making my mum a bit nervous by sitting in some open window. The problem was that the window was
30. all, but that’s normal and I have nothing against some order of things. What bothers me is the waking up
31. . Her job must be tiring… So I tried my luck with some other counter. In vain. Break again, just before
32. I really would like to put in life is the move to some other place, with vast meadows and moors... but s
33. urnalism, so the JOB would here mean I'd work for some paper or maybe in the radio; I was into that stuf
34. to be pleased with everything. As people noticed, some particular line has been crossed. Unfortunatelly.
35. ith little children. I think I would also work in some private language school. After MA... Well, I will

SOME + SINGULAR COUNTABLE NOUN (continued)

36. t my roommate? The guy is a first year student at some private school and his days look something like t
37. yet another exam mother) but still has to show me some proof of that. What she does though, is having he
38. vania Goofy Juice, and get bored in cinema during some pulp movie. Even MPK has worked as if its schedul
39. pringtime coming..to be more positive :), to have some relax before learning for the tests and so on...
40. e fancy Matrix-like) fights; Bruce Lee just kicks some serious butt (and not only) :) 3. Leon - an emot
41. m friends. My first book in English was probably some short story (maybe 20-30 pages). I think they are
42. d greetings, and then the history teacher noticed some sign on her T-shirt (in the chest area, so to spe
43. n May 1, 2006!!! What is more he booked for us in some small hotel at which the bands are also going to
44. f three. Probably I was on a walk with my parents some sunny Sunday afternoon. This one was taken by my
45. l hint - THE smile). photos Okay, now let's do some task. Briefly: first photograph on the right (lad
46. ee at 10 am with my friend so as to have at least some taste of normal life;)I didn't want to take the b
47. ot just "appeared" a couple of months ago, out of some unexplained twist of hormones. No, sir. It has be
48. he above mentioned, positive memories I have also some unpleasant experience. I will remember my first l
49. bout three then and it was during our holidays or some weekend. I always had to run because my sister (
50. n Obama. I am convinced that he will indeed bring some wind of change to this country; the more peaceful

AMOUNT 15 hits, 5 erros (4 underdifferentiation errors, 1 colligation error, 1 article omission error)

1. ns to be my sister's Name Day and so an additional amount of craziness follows. It's only noon and already
2. chapters I realized that the book carries a large amount of emotions, tells a story of revenge, and 'revea
   (colligation - amount of emotion?)
3. school the other evening and became stunned by the amount of hair that I’ve lost during the last 3 years. I
4. hat I have trouble coping with is the overwhelming amount of love problems those guys have. I've just learn
5. e status description: ... You should've seen the amount of messages I got over the following night. From
6. n't want to think about surviving a month with the amount of money I have. I want to be able to save some m
7. he can afford a fancy, a new Saab and an infinite amount of petrol, not to mention accomodation in hotels.
8. e that this profession might give you considerable amount of satisfaction. There are so many institutions y
   (missing article)
9. it to boil down only to a task:), there is a huge amount of sites which, depending on my mood and needs, a
10. emplate and I must admit that it cost me a serious amount of stress. Even though I knew the theory, how I s
11. matter of time;-(Have you ever thought about the amount of time wasted? So many things one could do, so m
12. I need my west and welaxation, OK? And there's no amount of your frowning upon it that is going to change
13. t a store/fast food bar and talking using enormous amounts of bad language. If you try to see beyond the ab
14. that Poland loses the trial and pays astronomical amounts of money from the state budget (read from people
15. shit before I saw it just on the basis of enormous amounts of Polish actors and directors saying just how g

It/This is|was (errors only)

1. hile it lasts because come June it could be over. It has made her stronger and more independent, since al
2. ioned perhaps? You definitely should not be! Why? It is mainly due to the fact that people all the time o
3. n your life no matter how the whole 'event' ends. It is vitaly important in order to reach, at least the
4. e date is that it is also Maja's dad birthday! :) It is the most precious gift he could get, I believe! M
5. le and went on a short trip (estimated 20-25kms). It is what I like: steady beat pupming in my ears, 2 wh
6. may judge the behavior, but not the whole person. It is not easy, at least for me. Second, I try to treat
7. I hope you do not mind Paweł writing on my blog. It is because I do not have an access to the Internet a
8. blog? I get some verbal feedback, but that's all. It is very discouraging, I must say... A perfect morni
9. w things and that they like the way I teach them. It is an extraordinary feeling! I don't want to say tha
10. are. Mine result was that I am most like Izzie:P It is true in terms of crying all the time and being a
11. wonderful artists' worth recommending to you all. It is Elaine Paige and Taro Hakase. The first artist is
12. erspectives.Even my imagination is helpless here. It is because I don't like thinking about my future as

ECONOM* (ECONOMIC/ECONOMICAL/ECONOMY/ECONOMICS)

1. orld's most influential country in time of a huge economic crisis ? I regret to say but I seriously doubt
2. tics but I believe that how he will cope with the economic crisis will tell about whether Americans have c
3. ountry, rich in resources, and very promising for economic development. The benefits of keeping friendly r
4. pecially appreciated in the UK and USA where both economic liberalism and conservative approach to traditi
5. em -- don't have to be so torn between family and economic needs. * Tee n drivers buckling their sea
6. ange. However, if Obama doesn’t manage to improve economic situation in the USA his greatest supporters wi
7. through;D) my friends got invitations form Kyrgyz Economical University and I'm going there with them :-)
8. to do to change the view of Poland from a small, economically backward country to a country West think of
9. you probably know, I graduated from University of Economics over two years ago. I studied logistics there,
10. what to choose. I decided to apply to Academy of Economics, and till this day I'm grateful to them for de
11. . Obama stated that he would improve the American economy and he would take actions to fight with terroris
12. unneeded warfare and maybe come to terms with the economy because, how much most of us don’t like, the eco
13. nomy because, how much most of us don’t like, the economy of the US has an enormous influence on the rest
14. olish reality at the moment. Politics, education, economy, sport.. NO COMMENT:) However, there is one posi
15. teresting theories concerning modern politics and economy. hope not everyone of them is true... so, that's

DURING – questionable uses

1. is my mum! As you can see - the picture was taken during Christmas time! I love it! The more I look at th
2. so busy that I couldn't even talk to them... But during Christmas we have made up for the lost time. We h
3. grandmother lived in the country and that is why during Christmas we used to spent some time outside. Whe
4. dying - draw giant robots or angry jelly monsters during class - argue with my friend who would win if we
5. inability to write even the simplest of programs during class, but our great performance on tests. Well,
6. believe in my abilities, for creating great mood during classes and making them memorable. I won't forget
7. ring. On the other hand, thanks to what I learned during classes, I've formed my views on many issues . I
8. bathe in the garden:):) And remember: no studying during Easter!!! What a shame! I've forgotten my passwo
9. comes the belly :) Things that made me happy.. During Easter: * I went to my hometown & spent a fe
10. "real" first encounter with computers took place during IT classes at primary school, during which we pla
11. ven't done some of the tasks yet. But, as usually during IT classes, I will do my best and catch up. Back
12. e a lot of noise and produced lots of heat:)) was during IT classes. I was taught how to turn on, operate
13. arch for pictures of "cookies". That is why, even during IT classess we may entertain ourselves, at least
14. top I often wonder how much the world has changed during last 15-20 years. I spent my childhood without a
15. thers. I really enjoyed pupils' positive attitude during lessons in a primary school. However, when I watc
16. stly, I'm happy I can learn so many useful things during my IT classes since I used to be hopelessly behin
17. some teaching materials and resources which I use during my private lessons. That is why when I started to
18. I just hate to be disturbed while at home. Sure, during social situations I'm far from sitting in a corne
19. this song when I saw Mateusz Ziółko performing it during the 'Mam Talent' tv show. I've also became fascin
20. preoccupied with many extra-curricular activities during the afternoon, and, what is more, children living
21. the coming holiday and song that I LOVE listening during the Christmas spirit. This is a song played by Tr
22. rtant, I felt satisfied with what I managed to do during the class and I was happy that the aims of the le
23. f characters and adventures, and it saved my mood during the cold and unpleasant winter evenings.
24. came stunned by the amount of hair that I've lost during the last 3 years. It really made me feel the flow
25. free to copy and take the full advantage of them during the preparation for our grammar classes. The onl
26. Coffee is the answer ;) Having many things to do during the weekend I thought about Daniel's recent sugge
27. rtainly one of most popular places to hang around during the weekends. My feelings about Gniezno? Well, I
28. that a good teacher should pose for somebody else during their classes and play a role of a mentor. It doe
29. plan" and my students probably felt like privates during their first days in army:) Fortunately, my "mento
30. I remember very well that one of his first words during this class was: "Americans always cause problems"
31. d true aspects of war and how a young man behaved during this hard time. Although the film is quite long,
32. higher culture and drop by to the museum not just during this one night in year. :) not that "Little Drag
33. and I'm going to do nothing conneted with studies during this weekend. I know that I was supposed to wirt
34. is happy of getting older :). Did I change a bit during this year? be a little, but not very much. I don'
35. Santa's knee. I wish you many pleasant surprises during this year's Christmas. Enjoy your time with your
36. quite sad and wistful…I think about what happened during those past three years, what changed, who I've me

COMMA + THAT – noun clauses

1. guistic determinism, the proponents of which claim, that human cognition is limited by the language. The P
2. mp the hell out of nowhere and you know for a fact, that something is wrong. For me such signs would, in r
3. usy clearing away freshly cut grass. How fortunate, that I should be passing that way, and too, since I wa
4. ess is stronger than me this morning. The thing is, that in order to be good with children I need to be ps
5. ul thesis out there. One thing worries me, that is, that I have so little time... Creating a masterpiece t
6. ndustry out of this stagnant state. The problem is, that he only does films about this one single characte
7. resh mind and head full of ideas, but the thing is, that I do not believe in great changes. US is too big
8. both sides' arguments. And it doesn't have to mean, that I totally agree or disagree with any of them. C
9. er to show its flaws and tries to convey a message, that it's not the kind of faith that matters, just the
10. w how to resist. That is why, it happens to me now, that I escape. my music1 On the first video you can
11. e. Why? Maybe someone will get inspired, will see, that despite obstacles and rather deliberate lack of m
12. served as a "great teacher" model. Is it really so, that few students want to make friends with their teac
13. but also by how he said it. It was clearly visible, that he knew exactly what has to be done to revive USA
14. lowed to mix together sound samples. The catch was, that the samples could only be 2, 4 or 8 seconds long,

COMMA + THAT – comma splice

1. t had a lot of pretty pictures... I am SO careless, that's ME. I am sorry, I know this blog was supposed t
2. . If you don't believe me, watch "My Name Is Earl", that'll really get you believing in the karma stuff.:)
3. nglish (God knows why I had not done it earlier:P), that's why in March 2006 I decided to go our college.
4. it. Good is good, Evil is evil. Good defeats evil, that's what I've always believed. Now, Good gets its h
5. ning, you're GRINNING. That's ok, grinning is good, that means on SOME level, you're laughing. :) Alright
6. , yeah. And? That's Dragon Ball, you don't like it, that's your choice. Me and my PSP are waiting in nervo
7. President. I like to think of myself as a liberal, that's why I'm glad that democrat won the elections in
8. . Billy Kilson remembered us from the gig in March, that was a surprise! And we talked like for hours in t
9. t wonderful thesis out there. One thing worries me, that is, that I have so little time... Creating a mast
10. This photo was taken at one of familiy meetings, that's why I'm wearing this jewellery;) My sister can'
11. ies because of this. Therefore, a bypass is a must, that’s a given. What I see as a gross dereliction, how
12. ew place, new people, and gosh, I have intenet now, that's really amazing ;) It's been very sad- moving ou
13. d (Poznañ - Warszawa - Gdynia - Warszawa - Poznañ), that's more than 1431 km by train. AND it was snowing
14. I'm planning on watching is around the semi-finals, that’s when every team wants to get ahead and the viol
15. d with it. I smiled to myself. What was I thinking, that some zombies will start walking down the street o
16. onished by the spectacle when I first saw it on TV, that is why I was very pleased to learn that finally t
17. presents and I feel old. That is pretty depressing,that is why during X-Mas I am deeply sad and angry wit

COMMA + THAT – relative clauses

1. in with, I must say that there is ONE music entity, that I adore, profess and worship. The Iron Maiden. Go
2. ige box full of buttons and twinkling little lamps, that made a lot of noise and produced lots of heat:))
3. The place where many ordinary, yet special, people, that I had pleasure to meet, lived. The place where I
4. all those delicious cookies,cakes and other sweets, that you know I love to bake. I have to do that becaus

COMMA + THAT – now that

1. ate German!!! I've recently discovered it... Now, that my exam is approaching real fast, I am positive-
2. was really surprised on the 4th of November. Now, that I've seen the presidential inauguration, I'm tota

HOWEVER – errors only

1. s 'Frankenstein'(Yeah!). It was hell of a choice, however after all the readers I went through earlier, I t
2. lready seen the photo during our grammar classes, however, now you have a chance to learn something more ab
3. y First communion I've bought (almost) brand new, however, MY OWN Amiga 500 for exactly 500 zlotys. It was
4. nificantly from their usual look, there might be, however, some additional performances, lectures or worksh
5. oss of memory (e.g. I forgot where I was running, however I still remembered about the new computer - so it
6. political situation of Poland? Hmm Hopefully yes, however, we have to wait some time in order to get a clea
7. I wish I could share with you more songs of hers, however there are too many of them. If you become interes
8. Maybe it is possible for me to achieve this goal, however certainly not in the five years' time. The more
9. ed permission to travel to the USA without visas, however, not Poland. Despite the fact that Poles supporte
10. f course the students sometimes became distracted however, that was normal short-term lost of concentration
11. h skills. I do enjoy teaching, the observations, however, are a different issue. Sometimes they are just m
12. as to make a nice photo of me in the city centre, however, as you can see I didn't really want to cooperate
13. turned out to be a really challenging experience, however this was just what I expected. My first lesson to
14. .. Love this song: 0:38 min gives me chills;) however, it's the lyrics of this song that I love the mos
15. August. There are more attractions of the town, however I usually meet with my friends there and discuss
16. w Year's Eve Hello everyone, Christmas is over, however we continue our holiday vacation by celebrating t
17. h. That is why I decided to put some more photos, however mayby without too many comments:))) The first one

HERE – word order errors

1. ;) But, there is one thing I love about my town - here time passes by very slowly and you don't have to
2. strict ;p, I would say that the thing that counts here the most is the brand of your car. Here you are a
3. that 'everybody knows everybody'. ;) People feel here 'like home' and, at the same time, can remain ano
4. ometimes read articles on this site. You can find here a lot of interesting and funny information about
5. As on many entertaining sites you will also find here some gossip about celebrities, lots of videos, ne
6. ,because it is a suburbian region and we have got here peace and quiet. I am really angry when somebody
7. tinks with fumes and dog excrements:D We have got here a lot of shops, especially Szarych Szeregów stree
8. hey have got imagination,I can tell;) We have got here even a shopping gallery,called "Galeria Podolany"
9. like her :) I almost forgot ;) here You have here official websites: Kelly Kelly 2 As far as my
10. sz's electro/hip hop, there's lot more electro in here than hip hop. Best party music in my opinion. Uff
11. things worth remembering, things that I've learnt here to tell! :) I know that I'm leaving this College
12. 4, but I'm not sure. My brother said that I look here like some kind of a 'little prince' but 'Dennis t
13. ow heavy metal sounds,enjoy. You have got lyrics here,just in case:) (...) my surroundings OK, now I
14. times... The past: I am more than happy to post here my old picture. I use it on my MSN and Skype, I l
15. log is a stupid idea either :)) I decided to post here some strange images I came across. From now on I
16. devoted to digital photography. I will be posting here some samples of my work. I'm barely an amateur :)
17. Values I've decided to put here some quotes and sayings about what I believe in.
18. ngers, Tori Amos. I know it would be wiser to put here a Tori's song instead of her photo, but I admit I
19. aaaaaada:D I was thinking about what song to put here that would remind me of our group and I've chosen
20. best song. They have so many of them. I will put here the one which most of you probably know - Wish yo
21. is a MIX:) But I have no idea who and WHY(?!) put here Paris Hilton..with this hamburger! : (Fragma - T
22. my cat IRA I adore her and now feel like putting here more of her pictures later;) See you later!!! T
23. n writing a diary. And the possibility of putting here short films, pictures, music and other 'fantastab
24. ng' lists, so this assignment is top notch.:) So, here goes, just like Karm, from 10 to 1, one being my
25. one :D Ok:) I think I'd better write something here about my favourite music;>. However it's too broa
26. ing my older brother. :P Probably, I am training here the ability of walking, but without a helping han
27. and was totally into journalism, so the JOB would here mean I'd work for some paper or maybe in the radi
28. my problems. I will, thus, try to share with you here some of my writings from the past two weeks, whic

NOT ONLY – various errors (labeled in brackets)

1. exactly me when I usually fall asleep. And it is not only the teddy bear which is similar to mine, but the (faulty cleft structure)
2. I've read the whole series about Harry Potter but not only. Today I just can't imagine NOT reading books in (confusing, incomplete structure, awkward sentence final position)
3. line:)) and I can say it was worth reading. It's not only the story about love to horses but also about the (not parallel)
4. line:)) and I can say it was worth reading. It's not only the story about love to horses but also about th (not parallel)
5. s more like a quick reference for translators and not only, in which you can find translations of words and (awkward clause final position)
6. ology and entertainment there. The website offers not only current news but foremost it’s a reliable source (blended structures)
7. around with more advanced programs. These made it not only possible to set the tempo, but also, or maybe most importantly, to compose my own (not parallel)
8. politics. So the change of American president is not only important for Americans but also for the rest of (not parallel)
9. ange will make good for the people of America and not only. Teaching practice Hmmm… Teaching practise is a (awkward sentence-final position)
10. Am I found this on Monika's blog. (...) Great! Not only doctors diagnosed slight OCD (Obsessive Compulsi (obligatory inversion)
11. give them that) I find those numbers disturbing. Not only there's more deaths of the good guys, but also m (obligatory inversion)
12. ;) was Commodore 64. I loved those simple games. Not only were they great fun, but also did kill time perf (awkward emphasis)

WH-words (when, where, what, why, how)- Word Order errors only

1. e severed, my death is sure. No one really knows when did riddles originate and where they come from. I
2. it was the year... 1988? or 1989 i don't remember when was i first allowed to play games on our wonderfu
3. ad but do we always have to stick to it? I wonder when was the last time when a kjo teacher conducted le
4. was a chubby tot at that time ;p I've got no idea where am I looking at in this photo. Maybe little Tinke
5. P It comes from 1990 so I was 4 years old:) Guess where am I? where's Wally?:P I know, I know, it doesn
6. e with all those wrong women... If somebody knows where is the perfect match for me please tell me. My f
7. city starting with a letter P (you probably know what are the two possible choices ;). I already have m
8. a and singing these songs but I couldn't remember what are these songs. Only after some time did I reali
9. and Alan Rickman. But stickin to the topic write what are your favourite movies. Here in comments or be
10. icant for understanding or I could't get any idea what could they mean from the context or just were int
11. I’m not sure about it. There were lot of question what does PO is going to do to change the view of Pola
12. so as you can see, at the age of 4 I already knew what does prostitute mean:D sorry Anita, but the word
13. as really taught time...I asked at least 5 people what should I do (priest included) .. I am glad of wha
14. tely, the teacher gave me some instructions about what should I do during the lesson)... Now, I think th
15. uy, really. I have no idea where he came from nor what was his name. I can't even recall whose wedding i
16. love this job;) To give u the slightest idea of what was it like in Krailling I'll put some pics, just
17. ious costumes on that day and when Santa asked me what was mine costume, I proudly replied: 'I'm a LADY!
18. of it...and it all seemed like magic!:) and now, what would be our life without a computer...
19. Obywatelska. I must say that I went there to see what’s he like and what is he going to talk about. Wel
20. physically. What can you tell when they ask you why don't you have anybody at the age of 24 and you c
21. he Hobbit] * Star Wars Trilogy [if you don't know why is it hre, you ain't JEDI ;)] * Fight Club [just
22. aid for that, or sth.. Anyway, and I asked myself why the heck do you have your blog? Why did you creat
23. I was , to whom I spoke and I wanted to cry...but why those people didn’t tell me I always tell people
24. o, we are to write about who forced us... I mean, why we decided to become students of the prestigious
25. re can I say,so just listen,watch and let me know how do you like her ;) her audition: "I surrender"
26. S won't mess it up in the end. If you are curious how does the new Windows work, you can download an im
27. n't remember whose song. Today I started thinking how is KJO? I don't have on my mind the building, of
28. sooo many great people around me :) I don't know how is that possible, but I have really GREAT LUCK wh
29. the men who I know would probably kill me now but how much I can stand - now it is the sixth hour when (Q!)
30. d unfortunately the desk folded:/....just imagine how shocking was the noise:)...and her comment was "i
31. ns will come, the list is infinite. I'm wondering how will they change the road leading to Lech stadium
32. it on YOU TUBE;-)Get a load of that! BTW I wonder how would you deal with the problem (after having 3 y

Appendix C: Pre-test

GRAMMAR III PRE-TEST name:_____

1. – Czy myślisz, że mama zaaprobuje (*approve*) wybór sukienki ślubnej mojej przyszłej bratowej?

2. – Żałuję, że nie miałem możliwości, żeby poznać lepiej twoją żonę.

3. Badania stwierdziły, że 70 % gospodyń domowych chciałoby podjąć pracę.

4. Czy możesz zaśpiewać do tego mikrofonu, który stoi bliżej fortepianu? (~~WHICH!~~)*

5. Podczas prac wykopaliskowych (*excavation*) stwierdziliśmy, że teren ten był zamieszkany od VII w.

6. Dzięki stronie internetowej mieliśmy możliwość zwiększenia ilości zamówień.

7. Geniusze potrafią w krótkim czasie nauczyć się niezwykłej ilości słów z języka obcego.

8. Ilość ludzi mądrych i wykształconych powinna stale wzrastać w nowoczesnej demokracji.

9. Kiedy właścicielka sklepu zorientowała się, że jakiś klient wybiegł z torbą i nie zapłacił, zadzwoniła na policję.

10. Komisja prawdopodobnie zatwierdzi nowy regulamin pomimo wielu zażaleń rozczarowanych mieszkańców.

11. Kto by pomyślał na przykład, że Lech zostanie mistrzem kraju?*

12. Lekarze walczą o życie rannego żołnierza, jednak stan jego jest krytyczny.

* Responses to these sentences were not included in the final data analysis.

13. Młodzież nie chce głosować na polityków, którzy aprobują (*approve*) nietolerancję, czy wręcz do niej zachęcają.

14. Musimy znaleźć jakiś sposób, żeby go ochronić przed tłumem nastoletnich fanów.

15. Najnowsza piosenka twojego zespołu została wykorzystana w jakimś filmie, prawda?

16. Nazwiska Havla czy Wałęsy są w świecie rozpoznawalne, aczkolwiek Gorbaczow jest bardziej znany.

17. Nie mam możliwości, żeby zmienić poprzednią decyzję kierownika.

18. Niektórzy lekarze nie rozumieją, że każdy pacjent potrzebuje życzliwości, jednakże sytuacja powoli się poprawia.

19. Poprosiłem, żeby polecił mi jakąś dobrą książkę o historii USA.

20. Stwierdzono w sondażu, że 12 proc. brytyjskich 14-latków pali papierosy.

    It...

21. Szkoła ma możliwości, żeby pomóc uczniom z ubogich rodzin, jeśli tylko dostanie pieniądze na ten cel.

22. To bardzo dobrze, że szef ceni cię jako pracownika. Musisz jednakże unikać przepracowania i stresu.

23. To zrozumiałe, że ilość państw zaangażowanych w wojnę z terroryzmem stopniowo spada.

24. Tylko niespełna połowa badanych (*respondents*) pochwala kontrowersyjną politykę Sarkozy'ego.

25. Wiele słów angielskich pochodzi z francuskiego, jednakże angielski to język germański.

26. Włamywacze zbiegli, pozostawiając znaczną ilość pieniędzy. Znaleziono też jakiś telefon komórkowy.

27. Rzecznik rządu stwierdził, że Ryszard C. jest oskarżony o morderstwo i dlatego nie zostanie wypuszczony z aresztu.

28. Tak długo jak istnieje możliwość uratowania ofiar katastrofy, akcja ratunkowa będzie kontynuowana.

29. Minister nie pochwala zachowania dyrektora w tym konflikcie, ale zdecydował się go nie zwolnić.

30. Trudno było przewidzieć ilość problemów związanych z budową tego teatru.

31. Poprzez analizę statystyczną stwierdzono, że duże gospodarstwa są mniej wydajne (*efficient*) niż małe.

Appendix D: Lesson reports and materials

This section contains lesson reports for all classes carried out as part of the study, as well as the materials used in them. The classes were paired in such a way that on a given day one group of students had one corpus-based class and one conventional class. This was not the case with the lesson on Saxon genitive, which took up a whole 90-minute session and was taught in both goups for the purposes of correlational analysis.

HOWEVER – corpus-based lesson

Experimental group – III/1+2

Students present: 20 out of: 21

Date: 12 December, 2010

Duration: 45 minutes

Course context (relevant preceding material): two classes on adverbial clauses

Objectives:

- students will be able to distinguish between concession (used in subordinate clauses, e.g *although*) and inter-sentence contrast (used with linking adverbs indicating contrast between separate sentences, e.g *however*);
- students students will be able to use the two types of links correctly in translation;
- students will be able to recognize errors involving contrast and concession.

Errors addressed: comma splice with HOWEVER (syntax), word choice (HOWEVER vs. ALTHOUGH)

Assumed L1 source of the error: The use of *jednakże, jednak, aczkolwiek,* etc. for both contrast and concession.

Materials used:

- concordance of HOWEVER from the BNC (15 lines)
- concordance of ALTHOUGH from the BNC (10 lines)
- concordance of HOWEVER from the blog corpus (15 lines)
- concordance of JEDNAKŻE from NKJP (12 lines)

Procedure:

1. Students analyze concordances of HOWEVER and ALTHOUGH, looking for regular differences in use;
2. Students are expected to observe regularities, and are guided to formulate rules of syntax and punctuation;
3. An error correction task is assigned, based on concordances from the blog corpus.
4. Students are given a series of concordance lines from the Polish corpus (NKJP) with the task to find a translation for the key word (JEDNAKŻE– a common

Polish equivalent of both ALTHOUGH and HOWEVER), which is a possible source of numerous errors observed in the data.

Comments: In the class preceding this one, it transpired that the concept of "concession" was unfamiliar to the students, so they found the distinction rather difficult to make. Only when doing the partial translation exercise did they seem to understand the problem.

Look at the examples from the British National Corpus below and compare the use of ***however*** and ***although*** for contrasting ideas.

they know they'll never be called upon to carry them out. If,	**however,**	you get elected and you then have to carry out your policies, then
unselfishly squaring the ball for Kiwomya to sidefoot into an empty net. Newcastle,	**however,**	refused to lie down and put Ipswich under heavy pressure in the last
ranging from videos to computers and home shopping, the survey adds.	**However,**	Gwynedd still has the highest percentage of active sports-people, despite a drop
to provide this. The feature stated that the Lieutenancy records have not survived;	**however,**	the Public Records Office at Kew does have the pay returns for the unit
user's point of view, can be safely left with the motor designer.	**However**	it is important to note the relationship between static torque and phase current when
was defamatory and, if so, what should the damages be. Now,	**however,**	having regard to what was learnt by C.N.L.'s counsel whilst sitting in court
understood by ordinary literate men and women ". Interpretation of the statute has,	**however,**	been technical. Since some terms such as " trust " are technical, a
to refer to one or more of the texts listed below. I would,	**however,**	sound a note of caution. It seems to me that the time,
with certain individuals or groups, will tend to effect the researcher's judgement.	**However,**	this is not to imply that the observer simply has to remain "detached
sometimes concrete --; sides, and duckboards. Allied trenches and dug-outs,	**however,**	never had the same air of permanence that many German ones exhibited,
were "tigers", preferring "blood and guts". For North,	**however,**	to be seen among the local people, to be seen to be doing
to collaborate and to get people who control resources to part with their cash.	**However,**	to ensure that projects do not collapse when the "entrepreneur's" work
want divorce and we want it now," they shout on a march.	**However,**	because the absurd law, Germi's principal target, was changed at the
you can include them in the same army if you wish to do so.	**However,**	you may only have one General, so you will not be able to
title only, whereas only 4% of the papers were read in full. Crane,	**however,**	argues 45 that affiliation to a prestigious university is more likely to lead to

. Others have been offered jobs on lower grades and salaries or short contracts.	**Although**	the redundancies aren't a direct result of the recession, they add to
a go. He could not do it if he were not the landlord,	**although**	in general you cannot, in the late twentieth century, expect high-minded landlords to
A police spokesman said all four people held had now been released without charge,	**although**	some had been bailed to report back to the police at a later date.
For example there were 367 recorded deaths of buzzards over a ten year period,	**although**	according to the RSPB the real figure is likely to be ten times that number.
, in theory, there were no steam trains. This proved to be correct	**although**	the specially built funicular railway is an interesting example of that kind of
existence by labouring. Mere labourers, however, had no security at all.	**Although**	the farmers frequently made considerable profits, wages did not rise to correspond
, and was not formally surrendered until the last Turkish garrison withdrew in 1867,	**although**	effective Turkish control had ceased some fifty years earlier. The expulsion of the
architecture at its plainest and most sober, very English in its understatement".	**Although**	long deserted, it still gives a superb picture of the power and ambition of
authoritarianism, intolerance and provincialism persist in the reigning power structures -	**although**	now under the name of nationalism, "communism's opposite". And,
the rhythm of the train in keeping with the rhythm of the poem.	**Although**	there are no apparent rhyming schemes there is a regularity in the number of

Find errors in students' use of HOWEVER below. Then compare your findings with the use of the word "JEDNAKŻE" in Polish.

1. an educator of an early primary class. Currently, **however**, I'm devoting my time to my lovely nearly 2-year-
2. as to make a nice photo of me in the city centre, **however**, as you can see I didn't really want to cooperate
3. dous effect the song had on me. The second artist **however** I've come to know through Caline Dion. One of her
4. Drczen", "Drdzen". The most well-known names are, **however**, "Drzeń" or "Drżeń". Contemporary name was based
5. first time. What I learned from this experience, **however**, and what I'm certainly going to incorporate in m
6. h skills. I do enjoy teaching, the observations, **however**, are a different issue. Sometimes they are just m
7. h. That is why I decided to put some more photos, **however** mayby without too many comments:))) The first one
8. lready seen the photo during our grammar classes, **however**, now you have a chance to learn something more ab
9. Maybe it is possible for me to achieve this goal, **however** certainly not in the five years' time. The more
10. more after they will have finished their studies (**however** long it may take). Talk about fair world. The se
11. my dad to study for exams. My favourite pastime, **however**, was discovering the contents of the kitchen cupb
12. ome samples of my work. I'm barely an amateur :), **however** photography has always been my passion. I started
13. point of view, I've made a few (minor) mistakes; **however**, the 'rest' (which was even more important)- was
14. ruary are not favourable for celebrating the day; **however**, I firmly believe that time for Valentine's Day i
15. s 'Frankenstein'(Yeah!). It was hell of a choice, **however** after all the readers I went through earlier, I t

1.	Nie wykluczamy takiej ustawy w przyszłości,	**jednakże**	to musi być głęboko przestudiowane i policzone.
2.	nego kongresu w Polsce. Jej członkowie uczestniczyli	**jednakże**	w kilku tego typu przedsięwzięciach. W 1927
3.	zauważyć, że nowe rozwiązania zawierają sankcje,	**jednakże**	znacznie łagodniejsze od dotychczasowych, zawartych w
4.	Pilsku, że światu oko zbieleje! Dostali	**jednakże**	odpór ze wszystkich możliwych stron. Miano im
5.	duże zatory itd., itd. Pozostaje	**jednakże**	problem, którego rozpatrywana dzisiaj ustawa nie załatwia
6.	. Sejm wprawdzie przegłosował wykreślenie tej zmiany,	**jednakże**	skutki tego wykreślenia, związane z brakiem równoczesnej
7.	tempa zamalowuje gniew i frustracje, dostrzegając	**jednakże**	i pozytywne strony otaczającego go świata: Jestem
8.	zgłosiła wniosek o ogłoszenie upadłości tego banku,	**jednakże**	po kilkunastomiesięcznym okresie działań zmierzających do
9.	dokonywane w przyszłym roku. Komisja zwróciła	**jednakże**	uwagę, że zwiększenie podatku dochodowego od osób
10.	celowego skłócenia poszczególnych fragmentów, które	**jednakże**	, ujęte precyzyjnymi podziałami kompozycyjnymi tworzą
11.	ciąży na nas obowiązek dostosowawczy. Chciałbym	**jednakże**	, Wysoka Izbo, z uzasadnienia, które
12.	tym przypadku na początku też był szerszy,	**jednakże**	muszę powiedzieć, że jestem zadowolony z powodu

HOWEVER – conventional lesson

Control group – III/3

Students present: 17 out of: 19

Date: 12 December, 2010

Duration: 45 minutes

Course context: four classes on different types of dependent clauses (noun clause, relative clause, this class was followed by two classes on adverbial clauses.

Objectives:

- students will be able to distinguish between concession (used in subordinate clauses, e.g *although*) and inter-sentence contrast (used with linking adverbs indicating contrast between separate sentences, e.g *however*);
- students students will be able to use the two types of links correctly in translation;
- students will be able to recognize errors involving contrast and concession.

Errors addressed: comma splice with HOWEVER (syntax), word choice (HOWEVER vs. ALTHOUGH)

*Assumed L1 source of the error: T*he use of *jednakże, jednak, aczkolwiek,* etc. for both contrast and concession.

Materials used:

- a table showing all options in terms of sequencing and linking clauses for contrast or concession, including punctuation options;
- a sentence combining task;
- a transformation task

Procedure:

1. The teacher begins the class by reading example sentences from Polish, in which *jednak* and *jednakże* are used either as a subordinating conjunction or as a linking adverb. In each case students are to decide whether they hear one complex sentence (with a sub. conjunction) or two sentences (with a linking adverb).
2. Students are presented with a table which lists and defines the most important syntactic options concerning concession and contrast in English;

3. Examples from the table are analyzed closely and discussed, the teacher drawing attention to those cases which usually cause problems and are often source of error;

4. Students are given a sentence combining task, in which for each pair of sentences they are to use two different types of contrastive links – subordination (concession in a dependent clause) and linking adverb (in a separate sentence);

5. After the sentence completion task has been checked, students perform a transformation task, where the target sentence is to include a given linking device for concession/contrast.

Comments: The lesson proved relevant to students' needs, as many did not seem to distinguish clearly between the two different structures (subordination vs. adverbial link). Some students, however, gave the impression of seeing the language problem in question as one of little importance. (Quote: "Does it really make such a big difference?") The tasks did not prove very challenging as students seemed to perform them rather mechanically.

CLASS MATERIALS

Polish examples given at the beginning of the lesson:

a) Długo szukał w archiwach, jednak(że) nie odnalazł żadnych dokumentów swojego ojca ani jego rodzeństwa.
b) Trzy lata temu Marek Kondrat ogłosił, że kończy karierę filmową. Jednak złamał przysięgę daną samemu sobie i zagrał.
c) Pewnego dnia matka musiała wyjść z domu, jednakże surowo nakazała swoim pociechom, by nikomu nie otwierały, chyba, że zobaczą jej białą łapkę.
d) Tablice interaktywne okrzyknięto technologią, która „zrewolucjonizuje oświatę". Jednakże sama technologia nie dokona żadnej zmiany jakościowej bez udziału człowieka.

Contrasting ideas

Formula	Example	Explanation
(main statement), **but** (contrasting statement)	I'd really like to come to the party**, but** I have to study tonight.	**Use a comma or semi-colon (;) with 'but'. (*co-ord. conjunction*)**
(main statement), **in spite of** (contrasting idea) OR **In spite of** (contrasting idea), (main statement)	They continued on their journey, **in spite of** the pouring rain.	**Use 'in spite of' plus a noun, noun phrase or gerund (*prep.*)**
(main statement), **despite** (contrasting idea) OR **Despite** (contrasting idea), (main statement)	They continued on their journey, **despite Ø** the pouring rain.	**Use 'despite' plus a noun, noun phrase or gerund (*prep.*)**
(main statement), **although** (contrasting statement) OR **Although** (contrasting statement), (main statement)	We wanted to buy a sports car, **although** we knew that fast cars can be dangerous. **Although** we knew that fast cars can be dangerous, we wanted to buy a sports car.	**Use 'although', 'though', 'whereas', and 'while' before a subordinate finite clause. (*sub. conjunction*)**
(main statement) . **However**, (contrastive statement). OR (main statement) ; **however,** (contrastive statement). OR (main statement) . (Contrastive**, however,** statement). OR (main statement) ; (contrastive, **however,** statement)	Getting a good job is difficult**; however,** most people find one eventually. Getting a good job is difficult**. However,** most people find one eventually. Getting a good job is difficult**;** most people**, however,** find one eventually. Getting a good job is difficult**.** Most people**, however,** find one eventually.	**Use 'however'/'nevertheless' in an independent contrasting clause/sentence, after a semi-colon or a full stop. (*linking adverb*)**
(main statement); (contrasting statement)**, though**. *final position*	She says she'll reward me for my efforts; I don't think she will**, though**.	**Use 'though' at the end of the contrasting sentence (*linking adverb*)**

Combine the sentences below. Use HOWEVER or NEVERTHELESS (linking adverbs) and one other link from the table above (a subordinating conjunction or a preposition).

1. We'd love to stay for dinner. We have got to get going.
   We'd love to stay for dinner; ***however****, we have got to get going.*
   Although *we'd love to stay for dinner, we have got to get going.*
2. They decided to stay in the area. They had problems with the local residents.
3. Peter decided to visit India. He doesn't like exotic food.
4. There were a great number of people who came. The hotels were not equipped to handle them all.
5. The book is expensive. It's worth every penny.
6. It rained a lot when they went for holiday. They enjoyed themselves.
7. She smokes 20 cigarettes a day. She seems healthy.
8. I didn't got the job. I had all the necessary qualifications.
9. I said many silly things last night. I still love you.
10. Falling forecasts of energy demand make a new drive towards nuclear power unlikely. Nuclear research will continue.
11. Man does not live by words alone. He sometimes has to eat them.
12. Mike is not hard-working. He has managed to become a millionaire.
13. They're very rich. They still want more money.
14. The company is doing well. They aren't going to expand this year.

Transform the sentences using the words in brackets.

1. We won the match although our goalkeeper had been injured. (in spite)
2. Despite being old, my grandfather takes us all on holiday every year. (though)
3. Although it was dark outside, he had not drawn the curtains. (however)
4. My parents are away. I'm not throwing a party, though. (despite)
5. The authorities also plan to resettle promising farmers, whereas in the early 1980s it was the poorest who were given newly purchased land. (nevertheless)

POSSIBILITY – corpus-based lesson

Experimental group – III/3

Students present: 19 out of: 21

Date: 12 December, 2010

Duration: 35 minutes

Course context: revision of non-finite verb forms a month before, differences in use of gerund and infinitive discussed with noun clauses (3 weeks before).

Objectives:

- students will become more cautious about using the word *possibility* as an equivalent of Polish *możliwość,* which is a common source of error;
- students will understand the meaning of the word *possibility*, how it should and should NOT be used.
- students will be able to use a number of other words and phrases which are appropriate equivalents of Polish *możliwość* in various contexts.

Errors addressed: Use of *possibility* with meanings unconnected with likelihood and prospective events but with ability, competency and opportunity; use of *possibility* with an infinitive.

Assumed L1 source of the error: The most common errors involve the phrase "*to have a possibility to (do sth)" – a calque from Polish *mieć możliwość (żeby) coś zrobić*. Even though gerundial/nominal complementation is more common in Polish in this context (*mieć możliwość zrobienia czegoś*)[34], an infinitive is also possible and appears to cause a lot of confusion, especially in translation tasks where the source form in Polish has an infinitive.A clear case of underdifferentiation error: the Polish word *możliwość* has a much broader scope of meanings, covering several different words and phrases in English: *opportunity, ability, power* or *be able.*

Materials used:

- parallel corpus concordance of *możliwość* (lemma search) –first only the Polish part, then with English (10 fragments)
- BNC concordance of *possibility* (lemma search) – 17 lines

[34] See an NKJP search (PELCRA search engine): http://nkjp.uni.lodz.pl/?q=5uhyp9z

- BNC concordance of *able* (8 lines), *ability & capability* (9 lines), *opportunity* (9 lines) (lemma searches) –
- learner corpus concordance of *possibility* (13 lines)

Procedure:

1. students receive the Polish part of the parallel concordance and are instructed to translate the key word (forms of *możliwość*) with its nearest context into English;
2. the English part of the parallel concordance is handed out to students for them to compare their translations with what was used in officially published texts;
3. discrepancies and variants are discussed;
4. BNC concordances of *possiblity* as well as *able*, *ability*, *capability* and *opportunity* are distributed and analyzed; students are encouraged to observe differences in the meaning and use of the words; students' observations are elicited and their attention is drawn to the differences in grammar patterns (e.g., *possibility* is not used with the infinitive while the other words are, *possibility* often appears in existential clauses);
5. learner corpus concordance of *possibility* is analyzed for error recognition and correction.

Comments: The distinctions between the words seemed rather vague to students, perhaps because the choice was not a straightforward one: there were many of options to choose from. The parallel corpus task proved rather frustrating to students because their choice of equivalent in most cases was *possibility* and they felt. Overall, the lesson addressed the students' problems accurately, but was not very well received.

*Translate Polish phrases marked in **bold and underlined**.*

POL:	wiedziałam, że na miejscu matki nie było nikogo innego, tylko ona sama, ale że jej tożsamość właśnie, której nie mogła zastąpić żadna inna, znikła, a ja ***nie miałam żadnej możliwości*** spowodować, aby wróciła, nakazać, by zaczęła powracać.
ENG:	
POL:	***Granice możliwości*** w rzeczach moralnych nie są tak ciasne, jak myślimy: słabości to nasze, wady, przesądy, ścieśniają je.
ENG:	
POL:	Nigdy nie zapomnę, jak pośród otaczającego koszmaru, ***bez możliwości zobaczenia*** jakiejś przyjaznej twarzy, przyszła moja kolej na oglądanie fotografii — nigdy przedtem nie czułem się tak szczęśliwy.
ENG:	
POL:	Zauważyłem ją już wcześniej i, tak jak się spodziewałem, miałem dzięki niej upragnioną, choć mocno ograniczoną, ***możliwość obserwowania*** kilku metrów korytarza na zewnątrz.
ENG:	
POL:	Nie spodziewałem się już żadnej rewizji celi, chociaż tutaj ***taka możliwość zawsze istniała*** i dlatego musiałem postępować bardzo ostrożnie.
ENG:	
POL:	Wszyscy mają prawo do otrzymania skutecznej opieki sędziów i sądów w zakresie realizacji swych praw i prawnie uzasadnionych interesów. W żadnym wypadku ***nie można nikogo pozbawić możliwości*** dochodzenia swych praw.
ENG:	
POL:	Wszyscy wnoszą wkład w ponoszenie wydatków publicznych ***zgodnie ze swoimi możliwościami*** za pośrednictwem sprawiedliwego systemu podatkowego inspirowanego przez zasady równości i progresywności, który w żadnym wypadku nie może przybrać rozmiarów konfiskaty.
ENG:	
POL:	— ***Istnieje możliwość***, oczywiście ***tylko możliwość*** — rzekł Hargreaves — że przeciek powstał za granicą, a tutaj podrzucono dowody,
ENG:	
POL:	A potem myślę, jakby to było z Cynthia w Lourenço Marques. Mógłbym z nią naprawdę rozmawiać. ***Możliwość rozmowy*** o pracy dobrze robi na potencję.
ENG:	
POL:	— Nie marnuje pan energii — stwierdził Castle. — Ach, to pan, sir! Owszem. Pomagam rządowi ***w miarę swoich skromnych możliwości***, a poza tym niewielu miewam poważnych klientów po piątej.
ENG:	

In the English fragments below, find sequences which are equivalent to the Polish phrases marked in ***bold and underlined****.*

POL:	wiedziałam, że jej tożsamość właśnie, której nie mogła zastąpić żadna inna, znikła, a ja ***nie miałam żadnej możliwości*** spowodować, aby wróciła, nakazać, by zaczęła powracać.
ENG:	I knew that that identity irreplaceable by any other had disappeared and I was powerless to make it come back, make it start to come back.
POL:	***Granice możliwości*** w rzeczach moralnych nie są tak ciasne, jak myślimy: słabości to nasze, wady, przesądy, ścieśniają je.
ENG:	The bounds of possibility, in moral matters, are less narrow than we imagine: it is our weaknesses, our vices and our prejudices that confine them.
POL:	Nigdy nie zapomnę, jak pośród otaczającego koszmaru, ***bez możliwości zobaczenia*** jakiejś przyjaznej twarzy, przyszła moja kolej na oglądanie fotografii — nigdy przedtem nie czułem się tak szczęśliwy.
ENG:	I'll never forget it, sitting in the midst of a living nightmare without even a friendly face in sight and when it came to my turn to see the picture I looked at it and I never felt so happy in all my life.
POL:	Zauważyłem ją już wcześniej i, tak jak się spodziewałem, miałem dzięki niej upragnioną, choć mocno ograniczoną, ***możliwość obserwowania*** kilku metrów korytarza na zewnątrz.
ENG:	I'd noticed it earlier and, as I hoped, it afforded me a restricted but welcome view of a few yards of space on the outside corridor.
POL:	Nie spodziewałem się już żadnej rewizji celi, chociaż tutaj ***taka możliwość zawsze istniała*** i dlatego musiałem postępować bardzo ostrożnie.
ENG:	The chances of a cell search now were slender but the danger was always there so one had to be very careful.
POL:	Wszyscy mają prawo do otrzymania skutecznej opieki sędziów i sądów w zakresie realizacji swych praw i prawnie uzasadnionych interesów. W żadnym wypadku ***nie można nikogo pozbawić możliwości*** dochodzenia swych praw.
ENG:	All persons have the right to obtain effective protection from the judges and the courts in the exercise of their rights and legitimate interests, and in no case may there be a lack of defense.
POL:	Wszyscy wnoszą wkład w ponoszenie wydatków publicznych ***zgodnie ze swoimi możliwościami*** za pośrednictwem sprawiedliwego systemu podatkowego inspirowanego przez zasady równości.
ENG:	Everyone shall contribute to sustain public expenditure according to their economic capacity, through a fair tax system based on the principles of equality.
POL:	— ***Istnieje możliwość***, oczywiście ***tylko możliwość*** — rzekł Hargreaves — że przeciek powstał za granicą, a tutaj podrzucono dowody,
ENG:	" It's possible, of course, just possible, " C said, "that the leak came from abroad and that the evidence has been planted here.
POL:	A potem myślę, jakby to było z Cynthia w Lourenço Marques. Mógłbym z nią naprawdę rozmawiać. ***Możliwość rozmowy*** o pracy dobrze robi na potencję.
ENG:	and then I think how it would have been with Cynthia in Lourenço Marques. I could really talk to Cynthia. It helps John Thomas when you can talk a bit about your work.
POL:	— Ach, to pan, sir! Owszem. Pomagam rządowi ***w miarę swoich skromnych możliwości***, a poza tym niewielu miewam poważnych klientów po piątej.
ENG:	"Ah! It's you, sir. Yes, I do my little bit to help the Government, and anyway I don't get many real customers after five.

Try to recognize the most common patterns in which the key words appear. Then think of Polish equivalents for those words in each context.

originate on the planet of these putative alien space travellers? A third	**possibility**	connected with the idea that life on Earth originated from space is
in the cloudy future, for at a meeting in the Town Hall a new	**possibility**	emerged, bluntly headlined in the Standard as "Proposed Municipal Golf
evidence of teacher involvement in curriculum planning indicates little	**possibility**	for ordinary teachers. In Britain, where there has been a tradition of less
the workload of the average Member of Parliament. Home working is a	**possibility**	in some areas and the successful applicant may be eligible for a company
outcomes? If Microsoft is found to have a case to answer, the minimum	**possibility**	is that the company is forced to sign a consent decree under which it has
impact of reservation pressure in public libraries has been noted. The	**possibility**	of a title going "out of print" if selection decisions are delayed must also
, when you know you are truly loved and accepted, then there is the	**possibility**	of change and growth. Many couples enter into marriages with the highest
and systems should be designed in such a way as to minimize the	**possibility**	of error. Clear instructions are essential in indexing tools such as thesauri,
is open to anyone interested in the Medau method of teaching and the	**possibility**	of joining the Teachers' Training Course, either with the next intake in
the business community in Blackpool, who expressed concern about the	**possibility**	of local government reform? One thing in particular that frightens the
other suppliers are not readily available. The decision lies between the	**possibility**	of losing future business by stopping supplies or of running the risk of a bad
which promised to inject an air of greater freedom into Spanish life. The	**possibility**	of the CNT's evolving in a more moderate direction, if not a strong
a while in the third match at Headingley they were in contention, but the	**possibility**	of victory never remotely reappeared. The first casualty of the match was,
ticularly where inter-male competition is also involved. Nevertheless, the	**possibility**	remains that, as Wallace argued, many of the sex differences in plumage and
-Mellon University's Mach 3 microkernel; in addition, there is the	**possibility**	that IBM might use the NT kernel --; "we could use the NT microkernel
protests during the last century of colonial rule. There remains the	**possibility**	that much social protest was expressed indirectly through crime. In
was very great it was not felt to be unbounded. There was always the	**possibility**	that there might be other places, beyond the limits of the known, which

no psychologist would dare to make assumptions about normal human	**abilities**	by studying imprisoned young orphans! But even in the wild, such teaching
divorce proceedings. The father does not challenge the mother's parent-ing	**abilities**	or impugn her capacity as a good mother. But he wishes the children,
would n't suit, for whatever reason. "Not everyone has the drive,	**ability**	, energy or inclination to aim for the top --; and of course that goes for
-sideration when choosing DIP software. Customisation refers to the	**ability**	to tailor the software to your application and includes alterations, from re-
of subsequent non-observance. Arechaga had pointed out that the	**ability**	of the parties to confer a right on third parties subject to conditions is a legal
and raunchy again, he left me with enough self-esteem to believe in my	**capabilities**	as a single mother." Forty, more than any other landmark birthday,
developed to run under CICS on IBM mainframes. It provides the	**capability**	to implement CICS applications for various Unix environments without
such as bioengineering. The major economic power of IT is the	**capability**	to increase productivity of manufacturing and of certain service industries.
combines the characteristics of viewdata screens with the sort of search	**capability**	usually found on databases. Appian Way, for instance, offer data disks that

a square with a small gap in one of its sides. Normally subjects were	**able**	to detect the presence of this gap equally well whether the square was
unwilling years as an articled clerk in his father's law office, before being	**able**	to enrol as a student. The office appointment had been intended for his elder
park in Britain then this could be your lucky year. Last week we were	**able**	to exclusively reveal to Mirror readers a Warner Holidays offer that slashes £300
"It's wonderful," he said, "I haven't been	**able**	to give her jewellery for a very long time." David, of Aberdare
Psychology Department, of which more will be said later. It ought to be	**able**	to offer more psychological and psychiatric help to the men who inhabit its
flowers and foliage FIRST-TIME FRUITS There is nothing better than being	**able**	to pick a juicy apple or succulent pear straight from your own tree. Sue
get enough food through to stop people starving to death, you won't be	**able**	to stop those who are prepared to kill for the few precious supplies,"
the tax relief you gain if you give through a covenant). They were	**unable**	to find any link when other variables like prices and incomes were taken into

LASMO Pakistan team and a group from HMS Chatham provided an	**opportunity**	for contestants to display some unusual techniques. The bowling alley, which
irregular both in incidence and attendance, assemblies provided a golden	**opportunity**	for the dissemination of royal propaganda, to insist on the notion of "public
traditions of the game". I was surprised spectators hadn't been given the	**opportunity**	to assess their merits. The most well-known, Sticky Wicket, which won
or Personnel Department. SAYESO The Company offers employees the	**opportunity**	to buy Company stock by means of a Save As You Earn Contract --; a scheme
technology areas. Clubs and special interest groups will be given the	**opportunity**	to discuss technical matters on the third day. The conference dinner will be
If he is, attack while he is rising because then he will have little	**opportunity**	to launch a powerful counter-attack. Southpaw stance forces the opponent
In the 19 and 20 Centuries, many more artists welcomed the	**opportunity**	to move outside their usual medium and work with Wedgwood craftspeople.
was here for only one purpose, and now Luke had given her the ideal	**opportunity**	to pursue it. She took a sip of her coffee and said, &bquo;
players. I'm just annoyed that so many Liverpudlians will be denied the	**opportunity**	to support the lads and frighten the life out of the Russians. I really believe

Check the fragments below for errors in the use of the word 'possibility'. How would you correct them?

a Polish film "Zmruż Oczy" . However, there is a possibility that I might have changed it and adopted for
autofocusing, easy access to all it's functions, possibility of changing lenses, better image quality due
ell...aspirations...ambitions...desires...so many possibiliteies...and many limitations...many things that
st person. I was quite nervous.... Now I have no possibility to feel the Christmas atmosphere.....I miss
e my life without the computer. It gives a lot of possibilities especially when you have an access to the I
For young people Mogilno also offers a variety of possibilities to spend leasure time. Even though Mogilno
d helping her with baking as there were plenty of possibilities to sneak something. Then I remember that I
teacher in future. But if I don't have any other possibilities (not connected with tourism) I will take te
ave any plans to be a teacher. I was open to that possibility, though not sure. There was a moment when I t
But the best thing of all is the Internet and the possibility to keep in touch with my foreign friend?.
eally, comes easier than writing a diary. And the possibility of putting here short films, pictures, music
ormance in my earlier post. Can he win? Well, the possibility is always there, but I already has my favorit
Maybe it was just a matter of luck, but I had the possibility to work with a group of cheerful, involved an

POSSIBILITY – conventional lesson

Control group – III/1+2

Students present: 14 out of: 16

Date: 12 December, 2010

Duration: 35 minutes

Course context: revision of non-finite verb forms a month before, differences in use of gerund and infinitive discussed with noun clauses (3 weeks before).

Objectives:

- students will become more cautious about using the word POSSIBILITY as an equivalent of Polish *możliwość*, which is a common source of error;
- students will understand the meaning of the word *possibility*, how it is and is NOT used by native speakers of English;
- students will be able to use a number of other words and phrases which are appropriate equivalents of Polish *możliwość* in various contexts

Errors addressed: Use of *possibility* with meanings unconnected with likelihood and prospective events but with ability, competency and opportunity. The most common errors involve the phrase "*to have a possibility to (do sth)" – a calque from Polish *mieć możliwość (żeby) coś zrobić*. Even though a nominal form is more common in Polish in this context (*mieć możliwość zrobienia czegoś*), an infinitive is possible and appears to cause a lot of confusion, especially in translation tasks where the source form in Polish has an infinitive.

Assumed L1 source of the error: calque from Polish. A clear case of underdifferentiation error: the Polish word *możliwość* has a much broader scope of meanings, covering several different words in English, such as *opportunity*, *ability*, phrases with such words as *power* or *be able*.

Materials used:

- word maps of *possibility, opportunity, and ability* (source: http://www.visualthesaurus.com/);
- a list of example sentences with *possibility* and a few related words used with different meanings and in different structures (based on dictionary examples)

- a list of sentences in Polish to be translated into English (based on dictionary examples)

Procedure:

1. The lesson begins with the teacher's question:
   How would you say this in English?
   Nie mam możliwości spotkać się z tobą w sobotę.
   (I'm unable to meet you/It's impossible for me to meet you/I can't meet you) on Saturday.
   Expectedly, some students say **I don't have a possibility to meet you.*

The teacher gives feedback on the answers, and leads into the topic of the lesson.

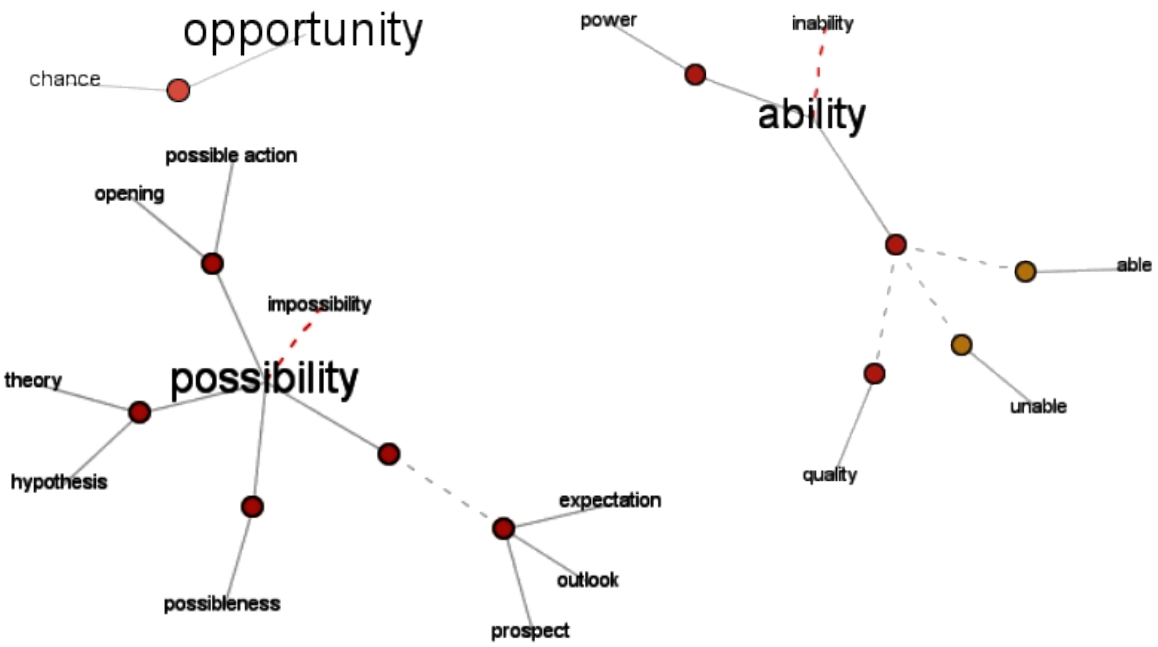
2. Students are given word maps of *possibility*, *opportunity* and *ability* to analyze. The meanings of words on the word map are discussed, so that students can distinguish between various uses and meanings of the key words. Then they look at the example sentences and are supposed to assign each of them to a particular area of the maps, depending on which meaning of *possibility/opportunity/ability* is involved. Students consult each other on their answers in pairs or small groups.

3. Students work on translating example sentences into Polish.

4. Feedback. The teacher suggests translations with the use of *possibility* if they have not occurred in the students' answers.

5. Students translate the Polish sentences into English, taking care not to misuse the word *possibility*. They are encouraged to use the other words discussed in class.

Comments: Using word maps is a novelty which is expected to make the lesson more effective and to neutralize the novelty effect in the experiment. This form of vocabulary presentation had been expected to be somewhat confusing to students, as it did not provide any context and so offered little information on how particular words were used. The example sentences were supposed to compensate for that to an extent, and it seems they did. Students were interested in the lesson, although some problems occurred when example sentences were being assigned to different areas on the word maps.



Which areas of the word maps above would you assign these sentences to?

1. This room has great possibilities.
2. There is a possibility that his sense of smell has been impaired.
3. Another possibility is that we'll go to Mexico instead.
4. There is a possibility that Britain will become a net car exporter once again.
5. This raises the possibility that a mechanism other than suppression is involved.
6. I will make the most of the chance to unwind in the Australian sunshine.
7. Maria will be a fine musician; she shows a lot of ability.
8. Patients are now able to buy the drug without a prescription.
9. The judge has the power to order a witness to give evidence.
10. Only the governor is now in a position to stop the execution.
11. It is not in her power to increase your salary, but she can recommend it.
12. Although we all thought the decision was unfair, we were powerless to change it.
13. All the children on board will have the opportunity to meet Santa.
14. We had so many guests that it wasn't possible to talk to everyone.
15. There was little chance of getting into such a defended area undetected.

How would you translate these sentences into Polish?

Translate into English:

1. Chciałbym zapytać, czy mają państwo możliwość wymiany mojej drukarki.
2. Istnieje możliwość, że jego słuch został uszkodzony.
3. Jeśli nas poprzecie, to nasza wygrana jest w granicach możliwości.
4. Kto ma możliwość zmiany tych przepisów?
5. Na egzaminie musisz być u szczytu swoich możliwości
6. Policja nie ma możliwości, żeby powstrzymać kibiców od awantur.
7. Powinniśmy rozważyć możliwość przeprowadzenia się do innego miasta.
8. Reżyser powinien umieć wykorzystać możliwości, jakie daje mu scenariusz.
9. Studenci nie mają możliwości korzystania z sieci uniwersyteckiej w tym budynku.
10. Te studia otwierają przed tobą wiele możliwości.
11. Nie dano mu możliwości obrony. Nie ma innej możliwości, jak tylko milczeć.

STATE – corpus lesson

Experimental group: III/1+2

Students present: 17 out of: 21

Date: 28 Februry, 2011

Duration: 40 minutes

Course context: First class after a semester-final test review; students are aware of their problems in translation, which is part of their end-of-year exam.

Objectives:

- students will become more cautious about using the verb *to state* as an equivalent of Polish *stwierdzić,* which is a common source of error;
- students will understand the meaning of the expression *to state that…*, how it is and is NOT used by native speakers of English;
- students will be able to use a number of other words and phrases which are appropriate equivalents of Polish *stwierdzić* in various contexts

Errors addressed: The most common error is the use of *state* when announcing or reporting research findings, as in *Researchers have stated that…* (for *Naukowcy stwierdzili, że…*), which in most cases is clearly inappropriate.

Assumed L1 source of the error: The straightforward source of the error may be the phonetic similarity between *state* and *stwierdzić*, as well as some overlap in use. One of many uses of *stwierdzić* may be translated as *state* – when the reported speech act involves some authority and refers to an official, formal announcement.

Materials used:

- Polish-English parallel corpus search results of *stwierdzić* (lemma search) – (10 fragments)
- a BNC concordance of stated (each line with a corpus section label)
- an NKJP concordance of stwierdzić (lemma search)

Procedure:

1. Students analyze the parallel corpus search results to find equivalents of *stwierdzić* in English. They are to underline the relevant fragments and are encouraged to voice any observations they have made. The teacher guides them to

the conclusion that rarely are stwierdzić and state used as equivalents. Other forms in English are pointed outinstead.

2. Students move on to the analysis of BNC concordance of *stated* in order to see in what ways the verb is actually used in English. The teacher draws students' attention to the corpus section labels, which are supposed to make students aware of the type of texts in which the verb typically appears.

3. Partial translation task: students consider the context of the key word in each line and decide how best to tranlate it in each case. The answers are first consulted in pairs and then with the whole class.

Comments: Students seemed genuinely surprised with the facts about the verb *state* which they were made aware of during the lesson. It seemed that they had became more comfortable with the corpus-based materials the second time they worked on them. Some students, however, did not respond well to the lesson. There were some awkward moments of silence during the lesson, during which students seemed to be waiting for the teacher to offer all the facts.

*Find and underline translations of forms of the word '**stwierdzić**' in the fragments below.*

POL: W przypadku gdy Komisja **stwierdzi**, że różnica między przepisami ustawowymi, wykonawczymi lub administracyjnymi Państw Członkowskich narusza warunki konkurencji na rynku wewnętrznym i powoduje w ten sposób zakłócenie, które powinno być wyeliminowane, podejmuje ona konsultacje z zainteresowanymi Państwami Członkowskimi.

ENG: Where the Commission finds that a difference between the provisions laid down by law, regulation or administrative action in Member States is distorting the conditions of competition in the internal market and that the resultant distortion needs to be eliminated, it shall consult the Member States concerned.

POL: Jeżeli **stwierdzi**, że doszło do naruszenia, proponuje środki właściwe do jego zaprzestania.

ENG: If it finds that there has been an infringement, it shall propose appropriate measures to bring it to an end.

POL: Jeżeli Rzecznik Praw Obywatelskich **stwierdzi** przypadek niewłaściwego administrowania, przekazuje sprawę do danej instytucji, organu, lub jednostki organizacyjnej, które mają trzy miesiące, aby poinformować go o swoim stanowisku.

ENG: Where the Ombudsman establishes an instance of maladministration, he or she shall refer the matter to the institution, body, office or agency concerned, which shall have a period of three months in which to inform him or her of its views.

POL: Było jeszcze orzeczenie lekarskie, ale nie mogło ono **stwierdzić** nic ponad to, że Harry zmarł, powiedzmy w granicach pół godziny, i w każdym razie orzeczenie lekarskie było warte tylko tyle, co słowa doktora Winklera: tego czystego, opanowanego człowieka, skrzypiącego pośród swoich krucyfiksów.

ENG: There was the medical evidence, but that could not prove more than that he had died, say, within a half-hour, and in any case the medical evidence was only as strong as Dr Winkler's word: that clean, controlled man creaking among his crucifixes.

POL: Zresztą Emma nie wydawała się już skłonna do słuchania jej rad; kiedy któregoś dnia pani Bovary ośmieliła się **stwierdzić**, że chlebodawcy powinni dbać o pobożność służby, odpowiedziała jej tak ostrym spojrzeniem i tak lodowatym uśmiechem, że poczciwa kobieta trzymała się odtąd z daleka.

ENG: Besides, Emma no longer seemed inclined to follow her advice; once even, Madame Bovary having thought fit to maintain that mistresses ought to keep an eye on the religion of their servants, she had answered with so angry a look and so cold a smile that the good woman did not interfere again.

POL: Wreszcie trójka ruszyła i wszyscy **stwierdzili** zgodnie, że doktor bynajmniej nie był uprzejmy. Uwagę zebranych odwrócił proboszcz Bournisien, niosący przez hale święte oleje.

ENG: At last the three horses started; and it was the general opinion that he had not shown himself at all obliging. Public attention was distracted by the appearance of Monsieur Bournisien, who was going across the market with the holy oil.

POL: Wysłali mnie do specjalisty, który **stwierdził**, że mam za niski poziom cukru we krwi, a nie za wysoki. Biedny Percival,

ENG: They sent me to a specialist who found I had too little sugar instead of too much ... Poor old Percival.

POL: **Stwierdził**, że przepada za piknikami nawet w upalne letnie dni, kiedy roi się od os i much, ale najbardziej lubi jesień.

ENG: He said that he loved picnics even on a hot summer's day when there were wasps and flies, but he much preferred the autumn.

POL: — Zostają jeszcze weekendy — **stwierdził** Castle bez przekonania. Zbyt dobrze pamiętał, jak wolno płynie czas w dzieciństwie.

ENG: "There'll always be weekends," Castle replied but without conviction. He remembered too well how slowly time limps by in childhood.

POL: W zakresie, w jakim umowy te nie są zgodne z Konstytucją, dane Państwo lub Państwa Członkowskie zastosują wszelkie właściwe środki w celu wyeliminowania **stwierdzonych** niezgodności.

ENG: To the extent that such agreements are not compatible with the Constitution, the Member State or States concerned shall take all appropriate steps to eliminate the incompatibilities established.

---

*Look at examples of '**stated**' from the BNC. Can you see how it is used? What is the style and type of text?*

1.	W_ac_polit_law_edu	desirous that it should be put into effect . The agreement	**stated**	that " in consideration of such desire " the executors would
2.	W_ac_polit_law_edu	not liable for the injury to the plaintiff . The court	**stated**	that " negligent conduct is more likely to break the chain of
3.	W_ac_polit_law_edu	the old paradigm is replaced with a new one . Kuhn	**stated**	the idea boldly , making paradigms very general and unitary and
4.	W_newsp_other_commerce	16(1) in the unamended form applicable at the relevant time	**stated**	: " Where a term of a contract purports to exclude or
5.	W_non_ac_humanities_arts	apprehend well-known criminals . In 1852 an official report	**stated**	that the " uselessness of the force , as at present constituted ,
6.	W_non_ac_polit_law_edu	re-entry to the international financial community . Fujimori	**stated**	, however , that the loan had been approved on July 29-30
7.	W_non_ac_polit_law_edu	use of foreign nationals as " human shields " , and	**stated**	emphatically that " the sovereignty of Kuwait is not negotiable
8.	W_non_ac_polit_law_edu	Party . The UN High Commission for Refugees (UNHCR)	**stated**	on Dec. 5 that more than 200,000 Kurds had fled their homes since
9.	W_non_ac_polit_law_edu	However , in an official joint statement Shar " and Genscher	**stated**	only that they had discussed " the situation in the Middle East
10.	W_non_ac_tech_engin	in the optimum position . However , for reasons previously	**stated**	these have not been used in the present design . The performance

How would you translate the key word in each example? (NKJP data)

1	. Sam uczestniczyłem we mszy w obrządku wschodnim i	**stwierdziłem**	, ze jest ona po prostu trochę inna.
2	założył Scenę Mimów przy Operze Kameralnej. Po kilku latach	**stwierdził**	, że aby się rozwinąć, musi być sam.
3	konsoli Xbox, która stała się hitem rynkowym. Ballmer	**stwierdził**	, że firma przez kilka lat intensywnie inwestowała w rozwój
4	małego Daniela miała trwać około półtorej godziny. Kiedy rodzice	**stwierdzili**	, że już jest za późno, wyszli z mieszkania.
5	wylegitymowany przez jordanowskich policjantów. Funkcjonariusze	**stwierdzili**	, że mężczyzna nie posiada wymaganego przy tego typu
6	. Szkoda jednak okazji. Oglądając biegi na 800 m	**stwierdziłem**	, że na pewno zakwalifikowałbym się do
7	ale kiedy się do nich zwróciłam, z przykrością	**stwierdzili**	, że nie mają ani jednego egzemplarza. Musiałam
8	". Wezwana karetka zabrała go do szpitala. Lekarze	**stwierdzili**	, że poparzeniu uległo 85 proc. jego ciała.
9	Wrócą czasy grzecznej telewizji? "Naukowcy	**stwierdzili**	, że prawdopodobieństwo zajścia w ciążę lub spłodzenia dzieci
10	taka drobna różnica. W czym? Otóż pan prezydent	**stwierdził**	, że Rywina zna: "Siwiec, Szymczycha czy
11	Olimpijskiego wywiesi flagę Unii Europejskiej. Uczestnicy spotkania	**stwierdzili**	, że sportowcy ze Starego Kontynentu już od dawna tworzą
12	- nie ma innych takich producentów w Polsce. Kontrolujący	**stwierdzili**	, że takie rozwiązanie wybrano na podstawie informacji o braku
13	Wymiotło sklepy! Byłem wczoraj w Berlinie i	**stwierdziłem**	, że zamknęli znane mi sklepy fotograficzne! Kiedyś
14	klasie A, jest sukcesem, zawodnicy po zakończonym turnieju	**stwierdzili**	, że zrobili więcej niż sami oczekiwali. : A
15	Gorbatko weszli na stację orbitalną w maskach tlenowych i nie	**stwierdzili**	aby znajdowały się w jej atmosferze znajdowały się niebezpieczne
16	Stacji Sanitarno-Epidemiologicznej w Wadowicach również nie	**stwierdzili**	błędów. - - mówi st. instruktor higieny zdrowotnej
17	nie zgodzili - dodała Krystyna Góźdź. Autor spornego artykułu	**stwierdził**	natomiast, że informacje na ten temat otrzymał w kuluarach
18	, która okazała się pomyślna dla Gromowskiej. Wojewoda	**stwierdził**	nieważność uchwały, o której dyskutowano na sesji 23 marca
19	że paliwa były celowo fał-szowane. W pobranych próbkach biegli	**stwierdzili**	obecność kilkuprocentową frakcji lekkich węglowodorów - typu
20	- denerwuje się Joanna. - W toku czynności sprawdzających	**stwierdziliśmy**	to, co znalazło się w uzasadnieniu o odmowie

STATE – conventional lesson

Control group: III/3

Students present: 14 out of: 19

Date: 28 February, 2011

Duration: 40 minutes

Course context: First class after a semester-final test review; students are aware of their problems in translation, which is part of their end-of-year exam.

Objectives:

- students will become more cautious about using the verb *to state* as an equivalent of Polish *stwierdzić*, which is a common source of error;
- students will understand the meaning of the expression *to state that…*, how it is and is NOT used by native speakers of English;
- students will be able to use a number of other words and phrases which are appropriate equivalents of Polish *stwierdzić* in various contexts

Errors addressed: The most common error is the use of *state* when announcing or reporting research findings, as in *Researchers have stated that…* (for *Naukowcy stwierdzili, że…*), which in most cases is clearly inappropriate.

Assumed L1 source of the error: The straightforward source of the error may be the phonetic similarity between *state* and *stwierdzić*, as well as some overlap in use. One of many uses of *stwierdzić* may be translated as *state* – when the reported speech act involves some authority and refers to an official, formal announcement.

Materials used:

- Examples of correct and incorrect sentences with the verb *to state* given by the teacher at the beginning of the lesson (spoken);
- Oral descriptions of a few situations in which the following reporting verbs are used: *state*, *find*, and *remark*;
- A list of ten noun phrases to be used in a indirect speech sentences with the verbs state, find and remark.
- A list of Polish sentences for translation into English.

Procedure:

1. Students listen to example sentences and need to decide whether the reporting verbs have been chosen appropriately.

2. Students listen to facts about a series of situations, for each of which a different reporting verb is used. Students are not allowed to take notes. Then they are supposed to reflect on the situations and recall the facts given about them from memory (a technique modeled on Gerngross et al. 2007).

3. Students build their own indirect speech sentences with ten noun phrases provided by the teacher. They are supposed to use *state*, *find* or *remark* in them. Care must be taken to make sure the verbs are used correctly by the students.

4. Students translate Polish sentences into English. Again, it is suggested to students that hey use *state*, *find* and *remark*. They need to decide which of the reporting verbs would be most suitable in each case. Later other verbs are considered as possible repalcements.

Comments: Students are slightly surprised that they are not allowed to take notes in the initial stages of the class. Once they realize how the task works, the class develops quite smoothly. The variety of forms offered in the lesson is smaller, but this may actually be the benefit of the lesson: students see the distinctions of function more clearly.

MATERIALS

Which is right and which is wrong?

**Researchers* have *stated* in a study that minor blows to the head add up over time and hamper the brain.

Vladimir Putin *stated that* Kazan should host all test events by June 2012.

Dr. Robertson unequivocally stated that he is against the use of illegal drugs.

*Harry stated that he didn't like going to the cinema on his own.

Spoken – in class:
Instructions: Listen to the description of a few situations. Try to remember as much as possible, but do not take notes.

The government spokesman *stated* that

- the elections would be held in October.
- there was no danger of another economic crisis.
- steps had been taken to evacuate tourists from Egypt.
- new cases of swine flu had been registered in the university clinic.

Do you think it was a particularly eventful day? Do you remember the news the spokesman announced?

Recent surveys have *found* that

- colleges are reporting increases in online enrollment.
- 28% of voters view gay marriage as a very important issue.
- there is a link between red meat and an increased risk of cancer.
- 93% of divorce cases were petitioned by wives.

Does any of the research results surprise you? Can you list them all again?

At the party people *remarked* that

- the food was excellent,
- the band played the kind of music they loved,

- they would like to be invited again,
- Britta was the most beautiful woman there.

Do you think it was a good party? What were the remarks people made?

Exercise 1

*Build 10 sentences with '**state'**, '**find'** or '**remark'** using the phrases below:*

- 75% of the 2000 adults surveyed believe
- a new recipe
- her new dress
- military presence
- nearly half the respondents
- non-profit organizations
- tax reductions
- the recent events
- those who...
- tickets for the concert

Exercise 2

Translate into English using 'state', 'find' or 'remark'. Think of other words you could use instead.

1. Badania stwierdziły, że kobiety i mężczyźni ogólnie rzecz biorąc mówią w tyle samo.
2. Gerhart Hauptmann stwierdził kiedyś, że tłumaczenia są jak kobiety – albo piękne albo wierne.
3. Prezes Nokii stwierdził, że gotowy jest nawiązać współpracę z firmą Google.
4. Paweł opowiedział rodzicom o Monice i stwierdził, że jest bardzo podobna do jego siostry.
5. Przedstawiciel pracowników stwierdził, że nie akceptują oni propozycji kierownictwa.
6. Stwierdziliśmy, że pogoda nie nadaje się na spacer i poszliśmy do kina.
7. Stwierdzono w eksperymencie, że witamina C nie ma żadnego działania w infekcjach wirusowych.

APPROVE – corpus lesson

Experimental group – III/3

Students present: 14 out of: 19

Date: 28 February, 2011

Duration: 30 minutes

Course context: First class after a semester-final test review; students are aware of their problems in translation, which is part of their end-of-year exam.

Objectives:

- students will become aware of the difference in meaning between *approve* and *approve of.*
- students will be able to use complementation of the verb approve correctly, depending on what meaning a given context requires

Errors addressed: Polish advanced learners of English tend to omit the preposition of after *approve* even though context indicates that it should be used.

Assumed L1 source of the error: Polish equivalents of *approve of* are simple transitive verbs which take nominal objects (*pochwalać coś, sprzyjać komuś, cenić coś*). Students are often unaware of the other meaning of *approve* (without the preposition), which is synonymous with *authorize* or *officially accept* (Polish: *zatwierdzić, zaakceptować*). The errors result from underdifferentiation between the two L2 forms, and the distinction is made more difficult by the use of equivalent words in L1.

Materials used:

- concordance of approve of and approve (without of), accompanied by frequency data on the two forms;
- learner corpus concordance of *approve*
- a gap filling task: concordance lines from which forms of *approve*/*approve of* have been removed.

Procedure:

1. Students analyze the BNC concordances for *approve* and *approve of*, trying to educe the differences in meaning between the two. Subsequently, students translate a few lines of each set into Polish to become aware of the source of the difficulty.

2. In the error correction task, students decide whether approve has been used correctly or not;

3. Students fill in gaps in blanked concordances, trying to deduce from context whether the preposition should be used or not.

Comments:
The lesson did not seem challenging and students did not have problems performing the task. Tests, however, did not show the issue to be totally resolved.

Analyze the concordances below. Try to define the difference between the two uses of 'approve'. Think of the Polish equivalent of 'approve' for each of the examples:

494 TOKENS in the BNC (*approve* + *of*)

1.	said . " The sort of films your mother would not	**approve of**	! " " Great . I wonder what my father would
2.	of me ? " Harry smiled . " Mrs Appleby will	**approve of**	anyone who has Kinmuire blood in their veins . " Aubrey
3.	now arouse political opposition and are be no means fully	**approved of**	by all the British people or even all British authorities .
4.	Spectacular by Joan Collins ! I am often asked whether I	**approve of**	cosmetic surgery -- and I do not . Too often ,
5.	Mr Hellyer said shortly . " I thought you did n't	**approve of**	manual labour , " Oliver said . Mr Hellyer straightened up
6.	no longer talk or laugh freely , as I knew he only	**approved of**	serious moods and studies . I fell under his freezing spell
7.	Thatcher 's wish for a moral revival in England . She	**approved of**	that , as evidently did the majority of her audiences ,
8.	never a favourite character of mine , as I do not	**approve of**	the British partition and subsequent desertion of India ,
9.	please him . " She was sure that Doc D would	**approve of**	the Caroline Charles electric blue evening suit (above ,
10.	should ask anybody in and around Nottingham whether they	**approve of**	the famous leaf stem statue , which costs £38,000 , the
11.	part in our surveys ; we hope you and your colleagues	**approve of**	the outcome . Our newly expanded page area came in handy
12.	person can still net $150 a week . Not every state	**approves of**	the telemarketing schemes , no matter how disguised they

5,241 TOKENS in the BNC (*approve* not followed by *of*)

1.	to allow an inspection would result in a course not being	**approved**	. Further permission for publication of the report was left to
2.	Eekelen , suggested a gradual approach whereby EC leaders would	**approve**	a common European defence policy at the Maastricht European
3.	on social , humanitarian and cultural issues unanimously	**approved**	a resolution condemning Myanma 's ruling military junta , the
4.	to finalise all the details with the design being personally	**approved**	by Her Majesty The Queen . Many different design styles were
5.	third round of voting . At the same meeting the Board	**approved**	medical relief on the application from relieving officers in
6.	Kourou , French Guiana [see p. 36811] , was	**approved**	overwhelmingly by the National Assembly on Aug. 8 . The Army
7.	June 27th 1988 suggests that a rule requiring the cabinet to	**approve**	sovereign investments was ignored . The letter was sent to Mr
8.	to 159 with eight abstentions . On Oct. 17 the Sejm	**approved**	Suchocka 's guidelines as presented on Oct. 9 by 166 votes to
9.	of Japanese troops to Cambodia The Japanese Cabinet on Sept. 8	**approved**	the dispatch of some 1,800 Self-Defence Force personnel to
10.	jointly as a " congress " in Versailles would have to	**approve**	the revision by a 60 per cent majority . Alternatively the
11.	to the council to put the job right as the council	**approved**	the work before they paid the contractor . " But Mr Buxton

Look at the fragments below. Find errors in the use of '*approve*':

1. public flattering in such form, I do not **approve** attacking with words as it has been done there, I
2. ents, and oh, how wonderful. And although I don't **approve** of Google's self-censoring for China, I think it
3. Especially in case of blogs. But I don't have to **approve** everything, and I do not have to be pleased with
4. lbums - to me that's enough. ;) Besides, I really **approve** of the fact that he does what he feels like doing
5. ng with words as it has been done there, I do not **approve** making people feeling sorry for what is not their
6. on't. Feel free to comment, if you wish. I do not **approve** public flattering in such form, I do not approve
7. ss. i understand to some degree, although i don't **approve** of, the reasons standing behind the ban on "prepa
8. think that he was an individual. I usually don't **approve** of delving into other people's lives and knowing
9. thought was the love of my life. My family never **approved** of him and when I left our contact broke complet

Fill in the gaps with either '*approve*' or '*approve of*' in appropriate forms:

1.	I can't help that . I don't		Georgina 's methods . That 's all there is to it
2.	established participated in this study . The study protocol was		by the Human Ethical Committee of the University of Basel and
3.	The latest survey of 1,473 church-goers shows 67pc		allowing the ordination of women , 22pc would vote against and
4.	membership of the Society is open to any person who		its aims . The subscription for full membership of the Society
5.	of its farmers . Budgetary issues On May 2 the Commission		its draft budget for 1992 -- the last under the five-year
6.	Overseas Development Institute -- A grant of £5,250 was		towards the cost of a suitable paper on condition that Christian
7.	The arrangement had not proved perfect . Aunt Lou seldom		the people who lived above her , but it worked after
8.	a few words with her , unless , of course , you		such behaviour . In which case , I shall begin making
9.	p. 38341] . Three days later the UN Security Council		a resolution , without a vote , which recommended membership for
10.	proposed by B H Payne , seconded by George Healey and		" unanimously and with enthusiasm . " William Blomefield
11.	" Lucy said evasively . " He doesn't much		what I 'm doing . " " Working backstage in a
12.	Secretaries of State and Defense had produced a paper which was		by the National Security Council : NSC 68 . In it was
13.	total of 95,000 Germans had their applications for citizenship		by the West German embassy in Bucharest , while the mayor of
14.	better than to ask . " But you can't possibly		drugs , " I said instead . " I neither approve
15.	waste a year . The government of the Bahamas has tentatively		a plan to allow a US company -- Summit Cement and Development
16.	what that meant . " It isn't every man who		women having ambition , " Ashley remarked a touch cryptically

APPROVE – conventional lesson

Experimental group: III/1+2

Students present: 17 out of: 21

Date: 28 February, 2011

Duration: 30 minutes

Course context: First class after a semester-final test review; students are aware of their problems in translation, which is part of their end-of-year exam.

Objectives:

- students will become aware of the difference in meaning between *approve* and *approve of.*
- students will be able to use complementation of the verb approve correctly, depending on what meaning a given context requires

Errors addressed: Polish advanced learners of English tend to omit the preposition of after *approve* even though context indicates that it should be used.

Assumed L1 source of the error: Polish equivalents of *approve of* are simple transitive verbs which take nominal objects (*pochwalać coś, sprzyjać komuś, cenić coś*). Students are often unaware of the other meaning of *approve* (without the preposition), which is synonymous with *authorize* or *officially accept* (Polish: *zatwierdzić, zaakceptować*). The errors result from underdifferentiation between the two L2 forms, and the distinction is made more difficult by the use of equivalent words in L1.

Materials used:

- Examples of correct and incorrect sentences with the verb *approve* given by the teacher at the beginning of the lesson (spoken);
- Oral descriptions of a few situations in which the verb *approve* is used in both variants (with and without *of*);
- A list of ten noun phrases to be used in sentences with either *approve* or *approve of*;
- A list of Polish sentences for translation into English.

Procedure:

1. Students listen to example sentences and need to decide whether the verb *approve* has been used appropriately.

2. Students listen to facts about a series of situations, for each of which a for of *approve* is used. Students are not allowed to take notes. Then they are supposed to reflect on the situations presented and recall the facts given about them from memory (a technique modeled on Gerngross et al. 2007).

3. Students build their own sentences with ten noun phrases provided by the teacher. They are supposed to use *approve / approve of* in them. Care must be taken to make sure the verb is used correctly by the students.

4. Students translate Polish sentences into English. Again, it is suggested to students that hey use *approve/approve of*. They need to decide which variant of verb complementation should be used in each case. Later other verbs are considered as possible repalcements.

Comments: The group seemed well involved in the class. Their interest was sustained by the challenge of remembering large chunks of information and eproducing it. The translation task proved a challenge in terms of lexis, but generally students had a sufficient amount of practice and did not find the problem itself very hard.

MATERIALS

Spoken Instructions: – (in class)

Which is right and which is wrong? What corrections should be made?

*Those who *approve* the death penalty claim that life imprisonment is not effective.

The White House has *approved* a new initiative in healthcare.

Unless the headmaster *approves* the new budget soon, teachers will protest.

*Some husbands do not *approve* the idea of their wives having a job.

Listen to the description of a few people and an institution. Try to remember as much as possible, but do not take notes.

Mr Right doesn't approve of

- women smoking in the street.
- same-sex marriage.
- stem-cell research being financed by the state.

What is he like? (He is very conservative.) What are his views again?

My grandmother approves of

- state-funded contraception.
- genetically modified food.
- separation of church and state.

What is she like? (She is very modern/progressive.) *Do you remember what her convictions are?*

The governor has approved

- a new design of the state's logo.
- his advisor's decision to employ a new PR specialist.
- the ban on hunting in state forests.

What was his day like? He has made a few important decisions today. Can you remember what they were?

The Parliament has refused to approve

- a law allowing the export of large capital.
- the program for privatization of state property.
- the construction of a new dam on the San.

What is the Parliament's attitude? (It's very cautious and conformist.) Can you repeat what the MPs failed to approve?

Build five sentences with 'approve+N' and five with 'approve of+N' sing the phrases below:

- military budget cuts
- increase in gas charges
- medals for war veterans
- contraception for teenagers
- a regulation that allows bikes to ride on pavements/sidewalks
- a proposal for a referendum
- cosmetic surgery
- organ transplantation
- new senator
- financial report

Translate into English using the word "approve".

1. Czy pochwalasz karanie dzieci poprzez zakaz wychodzenia z domu?
2. Komisja Europejska może *zatwierdzić* polski Program Rozwoju Obszarów Wiejskich najwcześniej w lipcu.
3. Minister zdrowia powiedziała, że nie pochwala zatrudniania lekarzy w kilku szpitalach naraz.
4. Opinia publiczna ogólnie aprobuje decyzje prezydenta Obamy.
5. Parlament Europejski zatwierdził wczoraj prawo ograniczające sprzedaż fałszywych leków.
6. Sędzia odmówił zatwierdzenia ugody pomiędzy bankiem a klientem.

A / SOME – corpus lesson

Experimental group: III/1+2

Students present: 18 out of: 21

Date: 4 April, 2011

Duration: 45 minutes

Course context: A class on problems within the Noun Phrase (number, gender, countability), two classes on articles, and one class on subject-verb agreement.

Objectives:

- students will be able to choose the appropriate determiner to mark indefiniteness of a singular (countable) noun
- students will be able to recognize and correct errors in the use of some in front of singular countable nouns
- students will recognize the difference in frequency and usage of Polish indefinite determiner "jakiś" and English "some"

Errors addressed: Polish students overuse the English determiner "some" in front of singular countable nouns as an equivalent of Polish "jakiś", where an indefinite article would be a better choice.

Assumed L1 source of the error: Different distribution of the two determiners (Polish: in front of both countable and uncountable and singular and plural nouns/ English: unmarked use – in front of uncountable or plural nouns; marked use – with countable singular nouns in lexical phrases or for emphasis, usually negative semantic prosody)

Materials used:

- BNC concordance (20 lines) of ***some +sing. noun***
- parallel (Polish-English) corpus concordance of ***jakiś*** (lemma search) – 9 examples with sing. nouns selected
- concordance of ***jakiś + sing. noun*** from NKJP (14 lines)
- concordance of ***some*** from the blog corpus
- a list of 10 most common combinations of ***some + sing. noun*** and ***a/an + sing. noun***

Procedure:

1. Students are given parallel corpus materials to look for equivalents of Polish phrases including ***jakiś*** in the parallel English text (most, though not all, contain the indefinite article). Attention is drawn to the most common choice – the indefinite article. The two cases of ***some*** are analyzed for any justification of the choice (one is part of a phrase – *for some reason*, the other strongly **emphasizes** indefiniteness).

2. Students analyze BNC concordance of *some + sing. noun* . They translate the key word in its nearest context to see that "some" is not always translated into ***jakiś***. Then they analyze the nouns in the phrases, to find that most of them are either uncountable or, if not, then part of a set phrase.

3. Partial translation: Polish (NKJP) concordance of ***[jakiś]+ sing. noun*** translated into English. Most examples require ***a/an*** to be used, rather than ***some*** (but also: some music, some time, some nervousness)

4. Error correction – student corpus concordance. Most of the examples include ***some*** used inadequately in front of a singular countable noun. Students are given one statistic to consider with this task:
   in the BNC "some book" appears 12 times, while "a book" appears 3017 times.

5. A list of most common ***some/a/an +sing. noun*** combinations is analyzed, to let students observe the differences in usage (some words appear in both lists – time, way – so differences in meaning and use must be pointed out)

Comments: Students seemed to know the basic rule and initially the class seemed too easy for them, but when translating from Polish, they still seemed tempted to use ***some*** too often.

Look at how the word 'some' is used in front of singular nouns. (BNC data). How would you translate the phrase in the middle into Polish?

than in 1990 --; but Ford still leads the shrinking market, and derives	**some comfort**	from not being able to make enough diesel engines to meet the demand.
can be highlighted. However, the size of the database resulted in	**some confusion**	as to the best way to proceed with the enquiry, and the next few
also very intelligent, highly imaginative, and easily my best student.	**Some day**	, with average luck, she would be a good writer. I remembered
best seed-bed for a crop of corn, as the ridges are bound, to , to cause	**some degree**	an unequal ripening of the seed. Again, as it
This is so fundamental a point that I will develop this in	**some detail**	. As already discussed on page 55, Mystery has nothing to do with
the land on which the pylons are to be built and it has to negotiate	**some form**	of agreement with the landowner," he said. "These negotiations are
communication with the depths in themselves so that they may have	**some idea**	of what that is like for other people, as well as helping them to be
as they would in the wild. Colour me bad Please could you give me	**some information**	on Disco Fish. I recently saw them at my local garden centre.
's no fun, not getting a wink at night --; you could end up	**some kind**	of insomniac or a nymphomaniac, whatever the correct term is! You
Woolfson: "The Football League Board considered your request at	**some length**	but are unable to accept your proposals. "The League has a duty
at the end, must have been composed for national celebration at	**some point**	during the time of the monarchy. It rejoices in the unrivalled might of
, and he looked at her. And then, it seemed to him,	**some power**	possessed her for she too hissed and struck out at him. As if she
its way to bend over backwards and, and help the, the applicants for	**some reason**	or another, presumably to get this thing through. And a that surprises
are many, the belief in the paramount value of these Greek plays as in	**some sense**	forming one of humanity's fundamental documents is always present.
some of its confidence. "I suppose it's possible she could be having	**some sort**	of psychological block. She wants to believe my father was a changed
and the tackle was being retrieved at speed to recast. It's worth giving	**some thought**	to the retrieve. It's very rare to have a perfect following wind
Punjabi very well, I think he wa he has been living in India for	**some time**	. So that was most popular, very busy person and er, since
what their subject "is all about". Most people feel that they have	**some understanding**	of the focus of disciplines like History or Physics, but are less familiar
smoothing the material with her fingers. "I'm glad I was of	**some use**	to you," she said, unable to keep an edge of tartness
that come up. It may be that er a child is being ill-treated in	**some way**	er by somebody, and we can help to decide what to do about

*In the English fragments below, find sequences which are equivalent to Polish phrases marked in **bold and underlined**.*
(parallel corpus data: http://korpus.hiztegia.org *)*

POL: Matka mówi i mówi o jawnej prostytucji i śmieje się z tego jeszcze dziecka ukrytego dotąd na placówkach, które nagle zjawia się w biały dzień i naraża swą reputację w mieście, zauważone i poznawane przez wszystkich, z tą miliarderską chińską szumowiną, z diamentem na palcu niby ***jakaś młoda bankierowa***. I płacze.

ENG: My mother rattles on. And she laughs at this child hidden till then in outposts up-country and suddenly emerging into the daylight and shacking up in front of everyone with this millionaire Chinese scum, with a diamond on her finger just as if she were a banker's wife. And she weeps.

POL: — Nie opuszczę cię. ***Jakaś myśl*** uspokoiła go. — Co prawda żmija nie ma dostatecznej ilości trucizny, aby ugryźć drugi raz...

ENG: "I shall not leave you." But a thought came to reassure him: "It is true that they have no more poison for a second bite."

POL: ***Z jakiegoś powodu*** mój człowiek stracił go z oczu w drodze między domami Coolera i Anny Schmidt; zameldował, że Martins kluczył wieloma ulicami, i obaj odnieśliśmy wrażenie, że umyślnie wymknął się śledzącemu go agentowi.

ENG: For some reason my man lost him between Cooler's and Anna Schmidt's flats; he reported that Martins had wandered widely, and the impression we both got was that he had deliberately thrown off his shadower.

POL: Myślę, że człowiek patrzący na koniec świata, na spadający samolot nie byłby nastrojony rozmownie, a dla Martinsa z pewnością nastąpił koniec ***jakiegoś świata***, świata beztroskiej przyjaźni, kultu bohatera, zaufania, które zaczęło się przed dwudziestu laty — na szkolnym korytarzu.

ENG: If one watched a world come to an end, a plane dive from its course, I don't suppose one would chatter, and a world for Martins had certainly come to an end, a world of easy friendship, hero-worship, confidence that had begun twenty years before — in a school corridor.

POL: W głębi duszy czekała ciągle na wielkie wydarzenie. Niczym tonący marynarze, przebiegała zrozpaczonym wzrokiem dookolną pustkę, wypatrując w oddali, we mgłach horyzontu, ***jakiegoś białego żagla***.

ENG: At the bottom of her heart, however, she was waiting for something to happen. Like shipwrecked sailors, she turned despairing eyes upon the solitude of her life, seeking afar off some white sail in the mists of the horizon.

POL: - Albo otruć ***jakiegoś chorego*** - ciągnął aptekarz. - Chcesz, widać, żebym stanął przed sądem, zasiadł na ławie oskarżonych, skończył na szafocie? Nie zauważyłeś, z jaką ja zawsze manipuluję ostrożnością, choć mam przecież bajeczną wprawę?

ENG: "Or poison a patient!" continued the druggist. "Do you want to see me in the prisoner's dock with criminals, in a court of justice? To see me dragged to the scaffold? Don't you know what care I take in managing things, although I am so thoroughly used to it?

POL: Wrócił po tygodniu, by się pochwalić, że dzięki usilnym staraniom odkrył w końcu niejakiego Langlois, który już ***od jakiegoś czasu*** patrzy łakomie na jej posiadłość, nie podając jednakowoż ceny. - Mniejsza o cenę! - wykrzyknęła.

ENG: He came back the following week and boasted of having, after much trouble, at last discovered a certain Langlois, who, for a long time, had had an eye on the property, but without mentioning his price. "Never mind the price!" she cried.

POL: Niektóre kobiety popadają w obłęd. Niektóre puszczane są kantem dla ***jakiejś młodszej, zachowującej milczenie służącej***. Puszczane kantem.

ENG: Some of them go mad. Some are deserted for a young maid who keeps her mouth shut. Ditched.

POL: Istotnie, skoro tylko rozchodzi się o jakiś fakt lub o jakieś prawo poszczególne w przedmiocie, który nie został unormowany za pomocą ***konwencji jakiejś*** ogólnej i uprzedniej, sprawa staje się sporna.

ENG: Indeed, as soon as a question of particular fact or right arises on a point not previously regulated by a general convention, the matter becomes contentious.

How would you translate the Polish phrase 'jakiś (+ noun)' in each of these fragments?

pracuje 24 godziny na dobę. Nie ma sztywnych godzin pracy. Kiedy jest	**jakaś awaria**	, siedzimy i do 2.00 w nocy. Mamy tak wszystko zautomatyzowane,
nią zaszczepił to jeszcze nie wiem. Lesław Korzeniak Słyszałem, że jest to	**jakaś bakteria**	znajdująca się w wodzie. Jeśli by było u nas jakieś zagrożenie nią to
powstać np. jakaś świetlica lub poczekalnia. Przydałaby się też	**jakaś kawiarnia**	czy bar dla podróżnych. Marcin Popek Brakuje tutaj sklepu
co się działo w Nakle. Wszyscy urzędnicy państwowi, jeśli zdarzy się	**jakaś tragedia**	czy skażenie środowiska rezygnują z urlopów i przyjeżdżają do ludzi,
radio, które szumiało w tle. – I już nie gadaj, tylko włącz	**jakąś muzykę**	. Nowak wstał i podszedł do regalika z płytami, do „swojego" regalika
gdy kietrzanie szarżowali na bramkę Petro. Może gdybyśmy wykorzystali	**jakąś sytuację**	, mecz potoczyłby się jeszcze inaczej - stwierdził Furlepa. Po przerwie
pod niego stary „Przekrój", akurat na zdjęciu Soni Ziemann i artykule	**jakiegoś Johna**	Custance, przeczytała: „Piszę ten tekst w pełni kryzysu maniakalnego".
niewiarygodnej harówki. I wtedy nagle się zakochała. Znalazła sobie	**jakiegoś kolegę**	z pracy, młodego i chętnego... było lato, słońce stało wysoko
koledzy nie wystraszyli się Anglików: - Panowie, doszukujecie się u nas	**jakiejś nerwowości**	, a jej nie było. Z tego co wiem, gdy koledzy wychodzili z
i oplata wzgórze wstęgą ciemnego asfaltu, wznosi się wyżej i wyżej, ku	**jakiejś niemożliwości**	, ku miejscu, z którego nie da się zawrócić – chyba że złamie się
z Jastrzębską do wysokości marketu Tesco. – Prace miały ruszyć już	**jakiś czas**	temu, ale niestety nic się na razie nie dzieje. Konary zaczęły odpadać
, portfel? – Tak, był w kieszeni spodni. Tylko trzydzieści złotych,	**jakiś kalendarzyk**	, karta miejska, odcinek emerytury. Ale jest też komórka. – Nowacki
a raczej na chodniki. Niedawno szłam do domu z zakupami i	**jakiś rowerzysta**	przejechał obok mnie. Trącił mnie kierownicą tak mocno, że powstał
Prokuratura ustala także, czy w tym czasie z naprzeciwka jechał	**jakiś samochód**	. Niezależnie od wyników pracy biegłego, nie ma wątpliwości, że 73-letni

In which of the examples below is the word "some" used inappropriately? NOTE: The British National Corpus - "some book" 12 hits "a book" 3017 hits

1. I want a dog or a cat. I have a huge need to hug **some** furry animal. Yesterday, my neighbour told me he
2. I'm preparing dinner. I love cooking, especially **some** unusual dishes. Happily, my husband likes eating
3. ike (I hope it happens soon). Today I'll chcecked **some** new route somewhere between Miłostowo and Antonin
4. ith little children. I think I would also work in **some** private language school. After MA... Well, I will
5. l hint - THE smile). photos Okay, now let's do **some** task. Briefly: first photograph on the right (lad
6. m friends. My first book in English was probably **some** short story (maybe 20-30 pages). I think they are
7. mmer, when I would have plenty of time to ponder **some** stuff, including why I waste time writing a blog
8. n May 1, 2006!!! What is more he booked for us in **some** small hotel at which the bands are also going to
9. n the wings of wind Recently I was going through **some** blog of a person I used to know. I found this sho
10. of a beloved person:) Kids? I would love to have **some** but it won't probably happen;) Last but not leas
11. ould feel its MESSAGE as well. Now, let's provide **some** explanation concerning all these unusual effects.
12. ouldn't bring myself to seeing the rest of it; at **some** point I believe I even dozed off. And I guess tha
13. pringtime coming..to be more positive :), to have **some** relax before learning for the tests and so on...
14. rridor or carrying this huge, heavy TV to show us **some** interesting report and often having trouble to fi
15. 't manage to spoil our time there. Yes, there was **some** rain, but on the whole we had lots of sunshine, m
16. telly :-) what to believe in... I've been doing **some** serious thinking in order to answer this question

17. to it: For now that's all, but if you want to see **some** Enya's clips - visit my blog later :) sad. Today

A / SOME – conventional lesson

Control group: III/3

Students present: 17 out of: 19

Date: 4 April, 2011

Duration: 45 minutes

Course context: A class on problems within the Noun Phrase (number, gender, countability), two classes on articles, and one class on subject-verb agreement.

Objectives:

- students will be able to choose the appropriate determiner to mark indefiniteness of a singular (countable) noun;
- students will be able to recognize and correct errors in the use of *some* in front of singular countable nouns;
- students will recognize the difference in frequency and usage of Polish indefinite determiner *jakiś* and English *some.*

Errors addressed: Polish students overuse the English determiner ***some*** in front of singular countable nouns as an equivalent of Polish "***jakiś***", where an indefinite article would be a better choice.

Assumed L1 source of the error: different distribution of the two determiners (Polish: in front of both countable and uncountable & singular and plural nouns/ English: unmarked use – in front of uncountable or plural nouns; marked use – with countable singular nouns in lexical phrases or for emphasis, usually negative semantic prosody)

Materials used:

- task sheets with lists of five Polish nouns,
- stories produced by teams of students in class

Procedure:

1. Students are divided into teams of 2 or 3 and task sheets are distributed. The task sheet instructs students to write stories with the nouns listed (in Polish), using the word ***jakiś*** in the adequate form in front of them.
2. Students pass on their Polish stories to another team, and receive others from a different team (clockwise).

3. Teams translate the stories into English and then pass them over to another team (random distribution).

4. Students are instructed to check if the sentences they have received are correct. The teacher elicits the rule and explains the differences in use.

5. Students check the stories and report the results. All the stories are then told to the class from memory, without reference to the written texts.

Comments: Some of the students engaged in the task quite creatively and produced interesting, amusing texts, while others made minmum effort and only after being strongly encouraged did they write a longer stretch of text. The language point was quite straightforward to students, and they had little difficulty using the determiners correctly.

MATERIALS

The nouns on the ask sheets have been selected with the following formula:

3 singular nouns

1 uncountable noun

1 plural noun (English equivalent – uncountable)

*Write a story in Polish which would include five words listed on your task sheet. Each of them must be preceded by an appropriate form of the word "**jakiś**".* 1. film 2. kibic 3. śmieci 4. niepokój 5. zaproszenie	*Write a story in Polish which would include five words listed on your task sheet. Each of them must be preceded by an appropriate form of the word "**jakiś**".* 1. alkohol 2. kłamstwo 3. komputer 4. pieniądze 5. adres
*Write a story in Polish which would include five words listed on your task sheet. Each of them must be preceded by an appropriate form of the word "**jakiś**".* 1. bzdury 2. bar 3. piosenka 4. pomoc 5. kierowca	*Write a story in Polish which would include five words listed on your task sheet. Each of them must be preceded by an appropriate form of the word "**jakiś**".* 1. czas 2. medal 3. plotki 4. uczeń 5. telefon
*Write a story in Polish which would include five words listed on your task sheet. Each of them must be preceded by an appropriate form of the word "**jakiś**".* 1. informacja 2. samochód 3. słowo 4. szkody 5. koc	*Write a story in Polish which would include five words listed on your task sheet. Each of them must be preceded by an appropriate form of the word "**jakiś**".* 1. książka 2. meble 3. problem 4. przyjaciel 5. smutek

AMOUNT / NUMBER – corpus lesson

Experimental group: III/3

Students present: 17 out of: 19

Date: 4 April, 2011

Duration: 30 minutes

Course context: A class on problems within the Noun Phrase (number, gender, countability), two classes on articles, and one class on subject-verb agreement.

Objectives:

- students will understand differences in use of the words AMOUNT and NUMBER and recognize what is the source of the confusion
- students will be able to use AMOUNT an NUMBER correctly and recognize errors involving the choice

Errors addressed: using AMOUNT instead of NUMBER with countable nouns

Assumed L1 source of the error: Polish word "ILOŚĆ", which is used with both countable and uncountable nouns ('LICZBA' being an alternative, but much less consistently used than NUMBER in English).

Materials used:

- a set of concordances for 'number of' and amount of' from the BNC (15 each)
- a set of concordances for 'ILOŚĆ' from NKJP (partial translation)
- a set of concordances for 'amount(s) of' from the blog corpus
- a list of top 10 collocates (limited to PLURAL nouns!) for 'number' and 'amount' [0,2] from the BNC (fine-tuning of the rule)

Procedure:

1. Analysis: students analyze BNC concordances for 'amount of' and 'number of' and arrive at the rule for choosing between the two
2. Partial translation: occurrences of the word 'ILOŚĆ' in NKJP concordances are translated into English
3. Error correction: students look for instances of incorrect use of the word 'amount' in the blog corpus data

Comments:

Compare the use of the words 'number' and 'amount' in the fragments below. What is the rule?

at any moment." Accident estimate: "I want you to estimate the	**number of**	accidents that you think actually occur at this junction. Use a scale from
beds are occupied by patients with HIV-related diseases and where the	**number of**	AIDS-associated orphans increases inexorably. Picture a teenage girl in
Softer " data were, however, gathered: interviews were held on a small	**number of**	cases, and a single case-example was further explored through the use of a
as the problems themselves. Some diagnoses are primarily economic. A	**number of**	economists have placed some of the blame for poor economic performance
of Infinite and finite Mind. Unlike Schleiermacher, Hegel had a large	**number of**	followers who sought to carry on from the point he had reached. They
. The average for the total practice population was 8.2 items a year. The	**number of**	items prescribed differed across the age bands (table I) with a small
to 1989 was (in round figures) 1,000, whereas in 1990 alone the	**number of**	new joint ventures reached 4,000 perhaps with the help of a revival of
the adjacent farm and, not surprisingly, was used to provide power for a	**number of**	other uses. There is a large, narrow belt wheel on the front
gun went off accidentally during a search of his home. In 1987 the	**number of**	people shot dead in Britain by the police using firearms was five. The
play in a theory of communication. It should be added that there are a	**number of**	philosophical problems with Grice's theory (see e.g. Schiffer, 1972),
age of either 16 or 23 years. There is an apparent decrease in the	**number of**	polyps affecting each family member, but this may not be a valid
is obviously already very close to the later rondo form. Here, between a	**number of**	statements of the same theme or section, contrasting sections are
proportion sharing at that time was only one/fifth. Moreover, the total	**number of**	titles was very much larger. The decline shown in Table 3.9 was
In appearance they resembled the larger cars supplied by Brush to a	**number of**	undertakings around that time, with all the usual distinctive Brush features,
, Lancaster, Newcastle, Sussex, and East Anglia, and have written a	**number of**	unpublished papers which I have given in seminars at other similar
Mortensen (1933) concluded that the characters of the disk are subject to a	**number of**	variations, especially occurrence of spines, differences in scaling, but that

TODAY. "We are also waiting for results which could confirm that a fair	**amount of**	alcohol was taken." Mr Tree was quizzed by detectives trying to
without being hard and unflattering. 3 Balance the liner with a generous	**amount of**	black mascara. The make-up artist's tip for applying it is to look
' . In our sample, in the south both have lost virtually the same	**amount of**	ground; yet the Lib Dems have won seats by the handful. Tactical
will change facts. The man you knew was a stranger to me and no	**amount of**	insisting will alter that." Her own pain made her more angry,
computer is running, a program like DesignaKnit needs RAM. This is the	**amount of**	memory that is needed for the computer to undertake all the different
and fuel vapour are mixed in suitable proportions: the throttle controls the	**amount of**	mixture passing through the carburettor, and the mixture lever its qualit
the tanks to the top with cool water to simulate the Monsoon. Although the	**amount of**	new water is minimal, compared to the real Monsoon, he still finds
violin , viola and cello, a goodly quantity of songs, and a fair	**amount of**	orchestral music, as well as two operas and three ballets. The booklet
, control or dominate, you could well be on cloud nine. An unusual	**amount of**	planetary activity now also relates to involvements with people from
: less steel per hulk. But a second more ominous result is a larger	**amount of**	plastic scrap, mashed up with car seats, brake fluids, coolants and
with any great alarm. Almost, Sara thought, one might detect a certain	**amount of**	satisfaction in her tones. Oh well, she never would understand Jenny's
announced that they have just paid a world record price for a rather small	**amount of**	tea. What this means to you or me is that if we happen
MP is known as "The Beast of Bolsover"? 81 What is the	**amount of**	the deposit parliamentary candidates must pay? (To within £50)
understand his unease. Here was a case which had taken up an inordinate	**amount of**	time for no result, and here was I, some remote big-wig from
over-watering. Brassicas like well-manured ground and a reasonable	**amount of**	watering. BELLS TO RING IN SPRING CHARLES LYTE ONE of the

How would you translate the Polish word 'ilość' in each of these fragments?

Nielsen, pobierają ryczałt za występ, bez względu na	**ilość**	zdobytych punktów. Może on wynosić nawet 10 tysięcy marek
Nowej Hucie i Podgórzu. Oba te urzędy obsługują znaczną	**ilość**	mieszkańców, a ponadto w tych rejonach mieszka więcej niż
Brzesko są niepokojące. Pierwsze półrocze bieżącego roku obfituje w	**ilość**	przestępstw popełnianych w stanie nietrzeźwym. Liczba zakłóceń porządku
na nich rękę, zamieszał pieniędzmi, ale chyba ta	**ilość**	forsy wydawała mu się śmiesznie mała i nic nie wziął
łezką w oku wspominają 1994 r. - Wtedy każdą	**ilość**	zebranych warzyw rolnicy sprzedawali na pniu niemal prosto z pola
. Miasto Szczawnica Frekwencja - 46,1%,	**ilość**	uprawnionych do głosowania - 5.569 osób, za
wymiaru normalnego czasu pracy. Z kolei podaż pracy oznacza	**ilość**	pracy oferowaną na rynku pracy do sprzedaży. Dotyczy ona
pedagoga niecodzienną ofertą. - Jeśli zamówiłabym określoną	**ilość**	podręczników po cenach hurtowych, mogłabym skorzystać z
. Niewykryty do dzisiaj przez Policję sprawca wylał dość znaczną	**ilość**	rtęci na prywatną posesję, skażając grunt. W związku
były zrzuty angielskie, ale bardzo znikome, i pewna	**ilość**	samolotów została strącona przez niemiecką artylerię przeciwlotniczą.
tego roku spośród 24 mln Niemców (liczba ta oznacza	**ilość**	przejazdów przez polską granicę), którzy przekroczyli Odrę,
minut spadła wczoraj w Miechowie po godz. 17 znaczna	**ilość**	deszczu. Krótki, ale gwałtowny opad spowodował istne spiętrzenie
. W jednej chwili te dziewczyny wbiły w pacjenta niezliczoną	**ilość**	igieł, coś z nim robiły, coś majstrowały pospiesznie
żyje ponad tysiąc różnych bakteri a to 268 to niewielka	**ilość**	o innych rzeczach którymi ludzie się "wymieniają" to
"17 razy 5" W ciągu ostatnich dwóch miesięcy	**ilość**	najgroźniejszych przestępstw popełnionych w powiecie olkuskim spadła o

Check the use of the word 'amount' in the fragments below. Correct if necessary.

1. s to be my sister's Name Day and so an additional amount of craziness follows. It's only noon and already
2. chapters I realized that the book carries a large amount of emotions, tells a story of revenge, and 'revea
3. chool the other evening and became stunned by the amount of hair that I've lost during the last 3 years. I
4. at I have trouble coping with is the overwhelming amount of love problems those guys have. I've just learn
5. status description: ... You should've seen the amount of messages I got over the following night. From
6. he can afford a fancy, a new Saab and an infinite amount of petrol, not to mention accomodation in hotels.
7. that this profession might give you considerable amount of satisfaction. There are so many institutions y
8. it to boil down only to a task:), there is a huge amount of sites which, depending on my mood and needs, a
9. mplate and I must admit that it cost me a serious amount of stress. Even though I knew the theory, how I s
10. matter of time;-(Have you ever thought about the amount of time wasted? So many things one could do, so m
11. I need my west and welaxation, OK? And there's no amount of your frowning upon it that is going to change
12. that Poland loses the trial and pays astronomical amounts of money from the state budget (read from people'
13. a store/fast food bar and talking using enormous amounts of bad language. If you try to see beyond the abo
14. hit before I saw it just on the basis of enormous amounts of Polish actors and directors saying just how gr

Instructions:
Look at the most common <u>plural</u> nouns that follow **number** *and* **amount***. Can you see a pattern?*

WORD 1 (W1): **NUMBER** (3.19)*

	WORD	**W 1**	**W 2**	**W1/ W2**
1	WAYS	319	0	638.0
2	REASONS	244	0	488.0
3	FACTORS	233	0	466.0
4	OCCASIONS	197	0	394.0
5	CASES	300	1	300.0
6	YEARS	583	2	291.5
7	DAYS	134	0	268.0
8	STEPS	118	0	236.0
9	ISSUES	114	0	228.0
10	MEMBERS	110	0	220.0
11	COUNTRIES	108	0	216.0
12	JOBS	107	0	214.0
13	STUDIES	102	0	204.0
14	PLACES	95	0	190.0
15	QUESTIONS	95	0	190.0
16	PROBLEMS	94	0	188.0
17	WOMEN	183	1	183.0
18	DEATHS	90	0	180.0
19	CHANGES	83	0	166.0
20	EMPLOYEES	79	0	158.0

WORD 2 (W2): **AMOUNT** (0.31)

	WORD	**W 2**	**W 1**	**W2/ W1**
1	DAMAGES	31	0	62.0
2	EARNINGS	5	0	10.0
3	FEES	5	0	10.0
4	FUNDS	9	1	9.0
5	WAGES	9	1	9.0
6	DIVIDENDS	4	0	8.0
7	EXPENSES	6	1	6.0
8	DRUGS	3	0	6.0
9	PESTICIDES	3	0	6.0
10	COSTS	5	1	5.0
11	EXPLOSIVES	2	0	4.0
12	DISTRIBUTIONS	2	0	4.0
13	DEBENTURES	2	0	4.0
14	CFCS	2	0	4.0
15	COMMUNICATIONS	2	0	4.0
16	INCREMENTS	2	0	4.0
17	MINERALS	2	0	4.0
18	PHOSPHOLIPIDS	2	0	4.0
19	RATES	2	0	4.0
20	VITAMINS	2	0	4.0

amount of people – 23 hits, including 16 in the spoken section of the BNC
number of people – 1172 hits, including 174 in the spoken section of the BNC

Lists generated by the following BYU BNC query:

COMPARE: number amount; COLLOCATES: [nn2*] 0.2
(plural nouns, two words to the right of ***number***/***amount***)

*The numbers represent the mutual ratio of the two words in the BNC, i.e. there are 3.19 tokens of "NUMBER" for every token of "AMOUNT", and 0.31 token of "AMOUNT" for every token of "NUMBER".

AMOUNT OF / NUMBER OF – conventional lesson

Experimental group: III/1+2

Students present: 18 out of: 21

Date: 4 April, 2011

Duration: 30 minutes

Course context: A class on problems within the Noun Phrase (number, gender, countability), two classes on articles, and one class on subject-verb agreement.

Objectives:

- students will understand differences in use of the words AMOUNT and NUMBER and recognize what is the source of the confusion
- students will be able to use AMOUNT an NUMBER correctly and recognize errors involving the choice

Errors addressed: using AMOUNT instead of NUMBER with countable nouns

Assumed L1 source of the error: Polish word "ILOŚĆ", which is used with both countable and uncountable nouns ('LICZBA' being an alternative, but much less consistently used than NUMBER in English).

Materials used:

- task sheets with lists of five Polish nouns,
- stories produced by teams of students in class

Procedure:

1. Students are divided into teams of 2 or 3 and task sheets are distributed. The task sheet instructs students to write stories with the nouns listed (in Polish), using the word *ilość* in the adequate form in front of them.
2. Students pass on their Polish stories to another team, and receive others from a different team (clockwise).
3. Teams translate the stories into English and then pass them over to another team (random distribution).
4. Students are instructed to check if the sentences they have received are correct. The teacher elicits the rule and explains the differences in use.
5. Students check the stories and report the results. All the stories are then told to the class from memory, without reference to the written texts.

MATERIALS

The nouns on the ask sheets were selected with the following formula:
3 plural countable nouns, 1 uncountable noun, 1 plural noun (English equivalent – uncountable)

*Write a story in Polish which would include the five words listed below. Each of them must be preceded by an appropriate form of the word „**ilość**”.* 1. filmy 2. kibice 3. praca 4. śmieci 5. ostrzeżenia	*Write a story in Polish which would include the five words listed below. Each of them must be preceded by an appropriate form of the word „**ilość**”.* 1. kłamstwa 2. komputery 3. alkohol 4. pieniądze 5. strony
*Write a story in Polish which would include the five words listed below. Each of them must be preceded by an appropriate form of the word „**ilość**”.* 1. bzdury 2. krzesła 3. hałas 4. płyty 5. nauczyciele	*Write a story in Polish which would include the five words listed below. Each of them must be preceded by an appropriate form of the word „**ilość**”.* 1. kwiaty 2. medale 3. plotki 4. uczniowie 5. czas
*Write a story in Polish which would include the five words listed below. Each of them must be preceded by an appropriate form of the word „**ilość**”.* 1. informacje 2. cierpliwość 3. samochody 4. słowa 5. szkody	*Write a story in Polish which would include the five words listed below. Each of them must be preceded by an appropriate form of the word „**ilość**”.* 1. książki 2. meble 3. problemy 4. przyjaciele 5. światło

SAXON GENITIVE – corpus-based lesson (both groups)

Group: III/1+2 and III/3

Students present: 35 (18+17) out of: 40 (21+19)

Date: 18 April

Duration: 90 minutes

Course context: The last class in a series of classes on problems of the noun phrase.

Objectives:

- Students will be aware of the differences in function between Saxon genitive, periphrastic genitive (*of*+noun) and compound nouns;
- Students will be able to choose the most appropriate possessive forms in a given context;
- Students will be able to use modifiers with Saxon genitive correctly.

Errors addressed:

- overuse of the *of*+noun construction
- problems with pre-modification of the head noun
- apostrophe omission/misplacement

Assumed L1 source of the error: Saxon genitive causes both intralingual and interlingual problems to Polish learners, which makes the structure quite challenging. Iterference may be seen in the fact that Polish does not have a dual system of forms for the possessive function (inflectional and phrasal). Also, the written form with the apostrophy is unusual for Polish learners, since Polish uses diacritics to mark phonetic variants, and not inflectional. Hence errors of apostrophy misplacement and omission. Finally, word order and article problems arise from the structure of the Polish NP:

książka Hemingwaya – Hemingway's book

*najlepsza książka Hemingwaya – *the best Hemingway's book* /

(target form: *Hemingway's best book*)

Materials used:

- a list of forms of Saxon genitive (with examples) (BB)
- a list of functions of Saxon genitive (with examples) (BB)

- BNC concordance of Saxon genitive, double Saxon genitive, compound nouns and *of*+noun
- a sentence combining task
- a list of the most frequent phrases with Saxon genitive in the BNC
- a conventional "provide the right form" exercise (homework)

Procedure:

(1) The teacher elicits forms of Saxon genitive (BB), then expands the list with special cases and more complex structures;

(2) The teacher elicits functions of Saxon genitive (BB);

(3) Students receive concordances of Saxon genitive and are supposed to recognize which function is performed by each item in the concordance; some flexibility with the functional categories may be allowed;

(4) Students analyze examples of compound nouns and try to define a general relationship between the first and the second element in a noun compound;

(5) Students analyze examples of noun phrases with *of*+noun, trying to see how they differ in function from Saxon genitive. Special attention is given to cases with human reference nouns followed by periphrastic genitive (*of*+noun). Particular stylistic features are pointed out in each case).

(6) Sentence combining task – the resulting sentences must contain a possessive form.

(7) Students receive lists of most common singular and plural phrases with the Saxon genitive (BNC data). Where possible, they are supposed to complete the phrases with pre-modifiers. Later, students choose 10 examples in which they are supposed to add an adjective that will modify the head noun.

(8) Practice/Homework assignment: sentence completion with a properly structured NP (Saxon genitive, *of*+noun, compound noun)

Comments: The complexity of the grammar problem is perhaps higher than in the other lessons of the project; that is why a whole class (90 minutes) was needed. Many factors are involved in the choice of the grammar form, and some aspects of Polish make those choices even more difficult for Polish learners. Students tend to choose *of*+noun because it allows them to avoid problems of pre-modification involved with the Saxon genitive. The lesson has made students realize that the two

structures are not interchangable, and they should be aware what the factors affecting the choice are.

MATERIALS

FORMS (elicitation > BB):

sing noun + *'s*	my father's house
sing noun -*s'(s)*	Dickens'(s) novels
(regular) plural noun +'	my parents' house
irregular plural + *'s*	the children's room
NP + *'s*	the man next door's wife (*group genitive*)

COMPLEX NPs – options and special cases

John and Mary's parents – How many people?

John's and Mary's parents – How many people?

the Secretary of State's visit, somebody else's song

my brother's neighbor's dog

a joke of my father's (double genitive – indefiniteness stressed)

FUNCTIONS (elicitation > blackboard):

- possession (sb HAS sth) – *Mike's car*
- human relationship – *my friend's mother*
- origin (authors, places) *the girl's story* (the story the girl told); London's parks
- subject (doer) – *the Queen's arrival* (The Queen arrived.) – (OF possible: *the arrival of the Queen*)
- object of an action – *the prisoner's release* (they released the prisoner) (rare! *OF* – 1st choice: *the release of the prisoner*)
- measurement (time, distance) – *two hours' drive*
- used by/produced for, organized by/for - *a women's magazine* – descriptive (note the use of articles here!)
- produced by/from –, ∅ *cow's milk*)
- lexical phrases (*for heaven's sake, to get your money's worth*, etc.)

Look at the concordances below and try to assign one of the functions listed on the blackboard to each of them. In some cases there may be an overlap, and two functions can apply.

1.	she has a vested interest in apartheid. If the	**husband's shirt**	has lost a button , he says , " Mary ,	
2.	our horses to a small rail and knocked on the	**priest's house**	door . A young , thin-faced man with brown hair ,	
3.	that can be generated. Intended audience LIFESPAN users.	**User's Guide**	to LIFESPAN ABLE This manual explains how to use	
4.	his mouth all dabbled with wine , to look into	**Adam's eyes**	. And Adam had looked into his , the wine singing	
5.	Reading Police Station. A cyclist has been taken to	**Oxford's John**	Radcliffe Hospital after an accident on the A423 ,	
6.	on the part of their coaching staff , the	**Republic's squad**	that played against us in the St Patrick's Day match	
7.	" Mr Waldegrave is married to Caroline , head of	**Leith's School**	of Food and Wine , which is providing the catering for	
8.	. The interview did not go as the interview with	**MacQuillan's mistress**	had gone and Rain was soon climbing ashore , silently	
9.	fully as possible. The irregularities of this kind in	**Braque's work**	of 1908 are much more conscious and deliberate. In	
10.	revolutionary change in society. A large portion of the	**world's population**	sees Marxism as an illuminating and liberating body	
11.	his meetings and his broadcast speech , for the next	**day's evening**	edition . At 8.30 next morning , the weather having	
12.	morbidity , and remains effective against relapses.	**CHAIRMAN'S REPORT**	TO THE BAR This report was delivered at the Annual	

parson and his wife and children used to go there for their holidays.	**Lewis 's father 's parish**	was on the outskirts of Manchester and the vicarage was
past six . . It's twenty past six .	**Oxfordshire's children's line**	is being relaunched this evening ; posters
cheek " says the gallant Goodwill) at the Langbaurgh	**council's chairman's ball**	on Friday . One of the Labour lads is
home in Vallum Court, Newcastle, earlier this week, to contact the	**city 's coroner 's office.**	Hotel sold:One of Scarborough's leading South Cliff hotels
still affecting her . Stella realises that although her	**sister's husband's death**	was extremely tragic it was all the same a "

1.	words than to remember six disconnected words. Word families and	dictionary work	. Best for individual and pair work. The children work on lists
2.	two different ways. Firstly, it meant access to library housekeeping	circulation files	designed primarily for staff use which could also serve as a
3.	tin . The deeper mines filled easily with water but, unlike	coal mines	, they were safe from roof collapse. B. The mine owners
4.	government in the six counties of the north was given to the	Stormont government	in Belfast. The Nationalists held majorities on some local bodies
5.	their picture power with local photographer Beth Davidson. The	picture project	was part of a district wide experiment to involve children with
6.	be awarded to the party aggrieved. The same applies today:	contract law	places few limitations on how parties may contract in the market
7.	inside the group. Also contained in the report are a cash	flow statement	and balance sheet. TSW. As a result of the loss
8.	fled their homes since mid-November in anticipation of an imminent	government attack	[see p. 38598]. The government on Dec. 6 described
9.	to what's going on in its own territory and remove the	planning obstacles	to what is actually happening and just to make sure that it

10.	the court shall not make an order for the possession of a	dwelling house	let under a secure tenancy on the grounds set out in Part
11.	. It was an unwelcome dowry. Eric Stonebanks, the first	Group Secretary	, in his published survey of the Bedford Group of Hospitals up
12.	the passage, with the result that when subjects are given a	recognition test	they will falsely identify the inference as having occurred in the
13.	stealthily, which may be very confusing. Hidden charges on	bank accounts	are not helpful to anyone trying to see what is happening to

1.	not for builders, then certainly for lovers. Four or five	sets of headlights	had blazed rudely into the neck of the lane, then turned
2.	crashed killing all the poor lads that were in her. Miley	Taylor of Deepdale	, who was in the Home Guard at the time, went
3.	, mentions the marriage of one of his sisters to Ragnvald,	father of Ulf	, and verse 19 of Austrfararvísur includes an unnamed brother of Ulf
4.	in less glamorous specialties such as old age psychiatry but in all	areas of medicine	. A broad and imaginative approach to this problem is needed;
5.	, can be used to correct the defect. Coach study seeks	cause of sickness	By Christine McGourty Technology Correspondent COACH travellers
6.	Ophelia drowns someone will jump on stage brandishing a	packet of Comfort	. There won't be product placement on stage but you will
7.	be shown by initials and so indicate that there is an unbroken	chain of title	. (4) Make a note of the purchase price,
8.	longer entitled to apply', he says. Tim Guldimann,	head of economics	and foreign policy in the Swiss science policy coordinating group, says
9.	Scheme where all employees are encouraged to sit three	levels of exams	. The first level is the bronze award which employees take after
10.	northern relief road and w-- attempting to increase the er the	flow of traffic	on that route. But even having said that, er it
11.	undeniable interest and importance of semantic and statistical	studies of language	, they appear to have no direct relevance to the problem of
12.	Lake of Dreams. If you went all the way across the	Lake of Dreams	you'd end up in the Lake of Death. I always
13.	forever? To answer this question we need to know the present	rate of expansion	of the universe and its present average density. If the density
14.	and be ready to act immediately. Beware of the subjective	attitudes of landlords/landladies	about your appearance, habits and origins. And don't bother
15.	, then make a lower offer. Most vendors allow for a	margin of negotiation	anyway. If your offer is acceptable then you proceed by applying
16.	spiralled around her, each constellation pricked out in delicate	shades of fragrance	. She melded into it, observing her own flight of colours

II. Use two sentences to make one. Follow the example.

a) Daniel has a dog. The dog is called Bobo.
Daniel's dog is called Bobo.

b) Peter has two sisters. Their names are Sarah and Nicky.

c) The children have new hamsters. The hamsters are behind the bookcase.

d) I have six cousins. Their surname is Parker.

e) We have a pet cat. Her name is Dolly.

f) Jess has two brothers. They're at university.

g) James has a beautiful girlfriend. She comes from Spain.

h) The twins have a nice new French teacher. He isn't very strict.

SAXON GENITIVE

1. Where possible, add an appropriate determiner or modifier (e.g. *a* day's work).
2. Find at least 10 examples where you could add an adjective before the last noun (e.g. the government's *social* policy)

	DETERMINER/MODIFIER	**NOUN+APOSTROPHE+S + NOUN**	N
1.		DAY'S WORK	243
2.		PEOPLE'S PARTY	232
3.		PEOPLE'S LIVES	227
4.		ARM'S LENGTH	221
5.		HEAVEN'S SAKE	199
6.		EARTH'S SURFACE	184
7.		FATHER'S DEATH	183
8.		WOMEN'S MOVEMENT	182
9.		YEAR'S EVE	168
10.		GOVERNMENT'S POLICY	163
11.		MIND'S EYE	161
12.		CHILDREN'S BOOKS	158
13.		WATER'S EDGE	145
14.		QUEEN'S SPEECH	143
15.		NIGHT'S SLEEP	142
16.		MAN'S FACE	129
17.		LION'S SHARE	122
18.		GOVERNMENT'S DECISION	120
		DOLL'S HOUSE	56

AmE	DOLL HOUSE	COCA 69 (BNC 1)
	DOLL'S HOUSE	COCA 38
	DOLLS' HOUSE	COCA 1

	DETERMINER/MODIFIER	**NOUN(-S)+APOSTROPHE + NOUN**	N
1.		YEARS' TIME	254
2.		YEARS' IMPRISONMENT	243
3.		MAGISTRATES' COURT	210
4.		WORKERS' PARTY	163
5.		YEARS' SERVICE	124
6.		MINERS' STRIKE	116
7.		WORKERS' UNION	115
8.		TAXPAYERS' MONEY	91
9.		CONSUMERS' ASSOCIATION	90
10.		STUDENTS' UNION	87
11.		WEEKS' TIME	86
12.		FARMERS' UNION	85
13.		MONTHS' TIME	83
14.		SHAREHOLDERS' FUNDS	83
15.		MONTHS' IMPRISONMENT	79
16.		DOLLS' HOUSE	66
17.		MINISTERS' MEETING	60
18.		WINNERS' CUP	60
19.		DAYS' NOTICE	59
20.		TEACHERS' PAY	56

a. The two words at the beginning of these sentences can be combined in one of three ways:

i. using the Saxon genitive

ii. using a phrase with 'of'

iii. as a plain sequence (compound noun).

Do whichever is required. The sequence may be the other way round from that given.

1. (hour, work) I must do another ..... this evening.
2. (education, diploma) Recognised teachers must have either a(n) ..... or a degree.
3. (moment, notice) The understudy had to take over at a .....
4. (hand, rail) No ..... was provided.
5. (party, politics) I find it difficult to get excited about .....
6. (doll, house) One of the Queen's ..... is on show to the public.
7. (summer, day) It was a beautiful .....
8. (wine, glass) What a pity I have broken that .....
9. (room, ceiling) The ..... was crossed with ancient beams.
10. (today, newspaper) I read something fascinating in .....
11. (break, tea) Most workers insist on having a .....
12. (hair, breadth) He escaped by a .....
13. (matter, root) The ..... is that he never had the slightest discipline instilled into him at home.
14. (worth, money) When you go out, you like to feel that you have got your .....
15. (ceiling, bedroom) I noticed yesterday that there was a big crack in my .....
16. (tray, tea) I bought her a ..... for her birthday.
17. (door, death) When I went to see him I really thought he looked as if he was at .....
18. (magazine, women) The sentiment in most ..... is so cloying that men find them unreadable.
19. (year, absence) He looked very different after his .....20. (meeting, committee) A ..... has been called for tomorrow after-noon.
21. (mouse, church) He's as poor as a ..... but he gives himself the airs of a Nabob.
22. (cold, head) It is not so much 'flu as a very severe .....
23. (story, fairy) The Sleeping Beauty is one of the most charming ..... I know.
24. (story, pirate) The ..... enthralled his listeners.
25. (salt, bath) With all these ..... I shall smell most exotic.
26. (service, dinner) She bought a very elegant .....
27. (journey, day) He lives about a ..... from here.
28. (licence, television) I must get my ..... renewed.
29. (state, mind) It is very difficult to determine his ..... at the time of the crime.
30. (year, time) 'That ..... thou may'st in me behold.'
31. (shirt, collar) His ..... was decidedly frayed.
32. (milk, glass) He drank the ..... straight off.
33. (wit, end) I really was at my ..... to know what to do.
34. (fare, prison) ..... is not very appetising.
35. (house, power) A big new ..... has been built there.
36. (property, slum) There is still a great deal of ..... in London.
37. (thought, second) He answered confidently after a ..... .
38. (car, seat) Most ..... are adjustable nowadays.
39. (day, holiday) The boys had an extra ..... .
40. (memory, lapse) During his recital the pianist had an unfortunate ..... .

(from Bywater 1982)

Appendix E: Test results

The tables below list individual students' error rates for each group of items in the pre-test and post-test, followed by the gain value. A separate listing is included for the delayed tests, with the same organization of data.

POST-TEST GAINS

GROUP 1+2

CORPUS-BASED

HOWEVER			
student	PRE	POST	GAIN
1b	2	0	2
1c	2	4	-2
1d	2	1	1
1e	2	0	2
1f	0	0	0
1g	1	0	1
1h	4	2	2
1j	0	0	0
1l	1	1	0
2a	3	2	1
2b	2	3	-1
2c	4	1	3
2d	0	0	0
2e	0	2	-2
2f	3	2	1
2g	2	3	-1
2h	0	0	0
2j	0	3	-3
2k	4	2	2

STATE			
student	PRE	POST	GAIN
1c	4	4	0
1d	4	0	4
1f	1	0	1
1g	3	1	2
1h	3	0	3
1j	2	2	0
2a	2	0	2
2b	3	0	3
2c	5	1	4
2e	2	0	2
2f	2	0	2
2g	2	0	2
2h	1	0	1
2j	2	1	1
2k	4	1	3
2l	2	0	2

GROUP 3

CONVENTIONAL

HOWEVER			
student	PRE	POST	GAIN
3a	2	1	1
3b	0	4	-4
3d	2	1	1
3e	1	2	-1
3g	4	1	3
3h	3	0	3
3j	2	0	2
3k	4	3	1
3m	3	3	0
3n	3	2	1
3o	0	0	0
3p	3	0	3
3r	4	0	4
3s	3	5	-2
3t	2	0	2
3u	3	3	0

STATE			
student	PRE	POST	GAIN
3a	3	0	3
3c	3	2	1
3d	0	1	-1
3g	4	1	3
3h	3	0	3
3k	4	1	3
3m	3	3	0
3n	0	1	-1
3o	5	1	4
3p	1	1	0
3r	2	1	1
3t	2	1	1
3u	3	1	2
2d	0	0	0

GROUP 1+2

CORPUS-BASED

SOME/A			
student	PRE	POST	GAIN
1b	1	1	0
1c	1	0	1
1d	2	0	2
1f	5	1	4
1g	2	0	2
1h	1	2	-1
1j	2	3	-1
1l	0	0	0
2a	0	0	0
2b	1	0	1
2e	0	0	0
2f	0	0	0
2g	0	0	0
2h	3	1	2
2j	3	3	0
2k	2	0	2
2l	5	1	4

GROUP 3

CONVENTIONAL

SOME/A			
student	PRE	POST	GAIN
3a	2	0	2
3c	0	0	0
3d	1	0	1
3e	1	1	0
3f	0	0	0
3g	0	0	0
3h	2	0	2
3k	2	1	1
3l	0	2	-2
3m	3	1	2
3n	0	1	-1
3o	5	4	1
3p	0	0	0
3r	4	0	4
3s	0	0	0
3t	4	0	4
3u	0	0	0

GROUP 1+2

CONVENTIONAL

POSSIBLITY			
student	PRE	POST	GAIN
1b	4	2	2
1c	2	0	2
1d	3	3	0
1e	2	1	1
1f	1	1	0
1g	4	2	2
1h	4	0	4
1j	2	1	1
1l	4	0	4
2a	2	1	1
2b	3	3	0
2c	3	0	3
2d	1	2	-1
2e	2	3	-1
2f	2	3	-1
2g	3	0	3
2h	2	0	2
2j	4	1	3
2k	3	1	2

APPROVE			
student	PRE	POST	GAIN
1c	4	0	4
1d	3	0	3
1f	2	0	2
1g	4	1	3
1h	5	0	5
1j	5	0	5
2a	1	0	1
2b	2	0	2
2c	4	0	4
2e	5	0	5
2f	3	0	3
2g	0	0	0
2h	0	0	0
2j	2	0	2
2k	3	2	1
2l	2	0	2

GROUP 3

CORPUS-BASED

POSSIBILITY			
student	PRE	POST	GAIN
3a	0	1	-1
3b	2	1	1
3d	2	3	-1
3e	2	1	1
3g	1	1	0
3h	2	1	1
3j	3	3	0
3k	3	4	-1
3m	2	2	0
3n	0	0	0
3o	3	3	0
3p	2	0	2
3r	1	4	-3
3s	1	1	0
3t	3	3	0
3u	2	2	0

APPROVE			
student	PRE	POST	GAIN
3a	1	0	1
3c	4	1	3
3d	4	0	4
3g	3	0	3
3h	4	0	4
3k	3	1	2
3m	4	0	4
3n	1	0	1
3o	5	1	4
3p	3	0	3
3r	1	0	1
3t	3	0	3
3u	4	0	4

GROUP 1+2

CONVENTIONAL

AMOUNT/NUMBER			
student	PRE	POST	GAIN
1b	1	0	1
1c	1	0	1
1d	0	0	0
1f	2	0	2
1g	4	1	3
1h	1	0	1
1j	1	1	0
1l	1	1	0
2a	3	1	2
2b	4	0	4
2e	1	0	1
2f	0	0	0
2g	4	1	3
2h	1	0	1
2j	0	2	-2
2k	4	1	3
2l	5	1	4

GROUP 3

CORPUS-BASED

AMOUNT/NUMBER			
student	PRE	POST	GAIN
3a	1	1	0
3c	1	0	1
3d	5	1	4
3e	4	4	0
3f	1	1	0
3g	1	0	1
3h	3	0	3
3k	1	0	1
3l	2	1	1
3m	1	0	1
3n	0	1	-1
3o	2	0	2
3p	0	0	0
3r	0	0	0
3s	1	0	1
3t	0	0	0
3u	3	1	2

DELAYED TEST GAINS

GROUP 1+2

CORPUS-BASED

HOWEVER			
student	PRE	delayed	GAIN
1b	2	0	2
1c	2	0	2
1d	2	0	2
1f	0	0	0
1g	1	0	1
1h	4	0	4
1j	0	0	0
1l	1	2	-1
2a	3	0	3
2b	2	1	1
2c	4	0	4
2e	0	0	0
2f	3	0	3
2g	2	0	2
2h	0	2	-2
2j	0	1	-1
2k	4	4	0

STATE			
student	PRE	delayed	GAIN
1c	4	2	2
1d	4	1	3
1f	1	0	1
1g	3	1	2
1h	3	3	0
1j	2	0	2
2a	2	2	0
2b	3	0	3
2e	2	2	0
2f	2	0	2
2g	2	0	2
2h	1	1	0
2j	2	1	1
2k	4	2	2
2l	2	0	2

GROUP 3

CONVENTIONAL

HOWEVER			
student	PRE	delayed	GAIN
3a	2	1	1
3d	2	0	2
3e	1	2	-1
3g	4	2	2
3j	2	1	1
3k	4	0	4
3n	3	0	3
3o	0	0	0
3p	3	1	2
3r	4	0	4
3s	3	1	2
3t	2	0	2
3u	3	0	3

STATE			
student	PRE	delayed	GAIN
3a	3	0	3
3c	3	0	3
3d	0	2	-2
3g	4	2	2
3h	3	0	3
3k	4	1	3
3m	3	2	1
3n	0	0	0
3o	5	3	2
3p	1	0	1
3r	2	0	2
3t	2	1	1
3u	3	0	3

GROUP 1

CONVENTIONAL

POSSIBLITY			
student	PRE	delayed	GAIN
1b	4	2	2
1c	2	3	-1
1d	3	3	0
1f	1	2	-1
1g	4	3	1
1h	4	1	3
1j	2	0	2
1l	4	1	3
2a	2	1	1
2b	3	4	-1
2c	3	2	1
2e	2	3	-1
2f	2	1	1
2g	3	1	2
2h	2	2	0
2j	4	2	2
2k	3	3	0

APPROVE			
student	PRE	delayed	GAIN
1c	4	0	4
1d	3	0	3
1f	2	0	2
1g	4	0	4
1h	5	0	5
1j	5	0	5
2a	1	0	1
2b	2	0	2
2e	5	0	5
2f	3	0	3
2g	0	0	0
2h	0	0	0
2j	2	0	2
2k	3	2	1
2l	2	0	2

GROUP 3

CORPUS-BASED

POSSIBILITY			
student	PRE	delayed	GAIN
3a	0	2	-2
3d	2	0	2
3e	2	3	-1
3g	1	2	-1
3j	3	3	0
3k	3	2	1
3n	0	0	0
3o	3	4	-1
3p	2	1	1
3r	1	1	0
3s	1	0	1
3t	3	3	0
3u	2	2	0

APPROVE			
student	PRE	delayed	GAIN
3a	1	0	1
3c	4	1	3
3d	4	2	2
3g	3	5	-2
3h	4	0	4
3k	3	2	1
3m	4	0	4
3n	1	0	1
3o	5	2	3
3p	3	0	3
3r	1	0	1
3t	3	5	-2
3u	4	0	4

Appendix F: The survey – the questionnaire and results

ANKIETA

IMIĘ I NAZWISKO: _____

Celem niniejszej ankiety jest zebranie informacji dotyczących zadań językowych opartych na danych korpusowych, które Państwo wykonywali na zajęciach z gramatyki. Ze względu na konstrukcję badań ankieta nie może być anonimowa, ale bardzo proszę o udzielenie szczerych i w pełni prawdziwych odpowiedzi. Głównym celem jest poznanie Państwa odczuć i opinii. Analiza wyników tej ankiety może być w przyszłości opublikowana, ale tożsamość respondentów nie będzie ujawniana. Uprzejmie proszę o poważne potraktowanie tego zadania i udzielenie odpowiedzi na wszystkie pytania.

CZĘŚĆ 1 **Informacja o respondencie.**

Zakreśl odpowiedź, która opisuje Cię najlepiej:

Czy studiujesz bądź ukończyłeś inny kierunek na UAM lub na innej uczelni?	tak	nie
Czy planujesz kontynuować naukę na studiach II stopnia?	tak	nie

Jak oceniasz swoją znajomość języka angielskiego w odniesieniu do tego, czego oczekiwałeś przed studiami w KJO? (Podkreśl właściwą odpowiedź.)

SŁABO PRZECIĘTNIE DOBRZE BARDZO DOBRZE

Jakie znasz inne języki obce? Wymień je poniżej i wskaż poziom biegłości dla każdego z nich, podkreślając odpowiedni opis:

..... PODSTAWOWY
NIŻEJ ŚREDNIO ZAAWANSOWANY
ŚREDNIO ZAAWANSOWANY
WYŻEJ ŚREDNIO ZAAWANSOWANY
ZAAWANSOWANY

..... PODSTAWOWY
NIŻEJ ŚREDNIO ZAAWANSOWANY
ŚREDNIO ZAAWANSOWANY
WYŻEJ ŚREDNIO ZAAWANSOWANY
ZAAWANSOWANY

..... PODSTAWOWY
NIŻEJ ŚREDNIO ZAAWANSOWANY
ŚREDNIO ZAAWANSOWANY
WYŻEJ ŚREDNIO ZAAWANSOWANY
ZAAWANSOWANY

Czy któregoś z powyższych języków zacząłeś się uczyć zanim rozpocząłeś naukę angielskiego? Jeśli tak jest, zakreśl ten język kółkiem na liście powyżej.

CZĘŚĆ 2

1. *Zaznacz krzyżykiem TAK lub NIE (ewentualnie NIE WIEM) przy słowach i wyrażeniach wymienionych poniżej, aby opisać jak oceniasz zadania oparte na danych z korpusów językowych (konkordancje, listy słów i wyrażeń, itp.).*

	TAK	NIE	NIE WIEM
nudne			
mylące			
przekonujące			
zbyt trudne			
interesujące			
skuteczne			
przytłaczające			
zmuszające do myślenia			
zbyt czasochłonne			
przejrzyste i zrozumiałe			

2. *Podczas zajęć z gramatyki kilka problemów prezentowanych było za pomocą materiałów przygotowanych na podstawie danych korpusowych (konkordancje, listy wyrażeń, itp.) Czy pamiętasz, jakie to były zagadnienia? Wymień te, które pamiętasz:*

……

3. *Do jakiego stopnia zgadzasz się z poniższymi twierdzeniami? Zaznacz odpowiednią kolumnę krzyżykiem.*

	stanowczo się nie zgadzam	nie zgadzam się	nie mam zdania	zgadzam się	zdecydowanie się zgadzam
Zadania oparte na danych korpusowych pomagają mi **zrozumieć** omawiane zagadnienia językowe lepiej niż inne typy zadań.					
Zadania oparte na danych korpusowych pomagają mi **zapamiętać** omawiane zagadnienia językowe lepiej niż inne typy zadań.					
Zadania oparte na danych korpusowych są bardziej **skuteczne** niż inne typy zadań, to znaczy pozwalają mi osiągnąć lepszą poprawność językową w zakresie omawianych zagadnień.					

4. *Jak oceniasz skuteczność niżej wymienionych typów działania na lekcji, gdzie pracuje się nad zagadnieniami szczególnie trudnymi dla polskich uczniów języka angielskiego? Swoją odpowiedź oprzyj na własnych doświadczeniach. (Dla każdego typu zadania postaw krzyżyk we właściwej kolumnie.)*

skuteczność / zadanie	żadna	ograni-czona	przecię-tna	wysoka	bardzo wysoka	nie wiem
analiza przykładów podanych przez nauczyciela (tablica, wydruk, prezentacja, itp)						
analiza konkordancji z korpusów rodzimych użytkowników języka (np. BNC)						
analiza korpusów równoległych (tekst w języku polskim i ten sam fragment w języku angielskim)						
częściowe tłumaczenie oparte na konkordancji						
tłumaczenie pojedynczych zdań podanych przez nauczyciela na j. ang.						
tłumaczenie dłuższych fragmentów na j. ang.						
rozpoznawanie i poprawianie błędów w konkordancji z korpusu uczniowskiego						
rozpoznawanie i poprawianie błędów w osobnych zdaniach podanych przez nauczyciela						
rozpoznawanie i poprawianie błędów w dłuższym fragmencie tekstu						
łączenie zdań prostych w złożone i wielokrotnie złożone						
przekształcanie zdań (transformacje)						
uzupełnianie luk w konkordancjach (słowo kluczowe)						
analizowanie map leksykalnych						
listy słów z danymi dotyczącymi częstości ich występowania						
budowanie własnych zdań z użyciem omawianej formy językowej						
budowanie własnych tekstów z użyciem omawianej formy językowej						

5. *Wybierz JEDNO z poniższych twierdzeń, które najlepiej oddaje Twoje odczucia wobec wykorzystywania materiałów korpusowych na lekcji języka obcego. Swój wybór zaznacz zakreślając odpowiednią literę kółkiem.*

   A. Z początku nie podobało mi się korzystanie z takich materiałów, lecz później zmieniłam/em zdanie.
   B. Z początku podobało mi się korzystanie z takich materiałów, lecz później zmieniłam/em zdanie.
   C. Od początku nie podobało mi się korzystanie z takich materiałów i nie zmieniłam/em zdania.
   D. Od początku podobało mi się korzystanie z takich materiałów i nie zmieniłam/em zdania.
   E. Żadne z powyższych (wyjaśnij krótko)

   .....

   .....

6. *Dla uzasadnienia opinii wyrażonej w poprzednim pytaniu, wylicz wady i/lub zalety korzystania z zadań korpusowych w klasie na zajęciach językowych.*

   Wady: _____

   _____

   _____

   _____

   *Zalety:*_____

   _____

   _____

   _____

7. *Gdybyś miał(a) zostać nauczycielem, czy próbował(a)byś czasem włączać w swoje lekcje zadania oparte na materiałach korpusowych? Zakreśl kółkiem i/lub uzupełnij.*

   TAK NIE TYLKO z pewnym rodzajem uczniów (jakim?)

   .....

8. *Czy samodzielnie korzystasz z korpusu, aby sprawdzić użycie angielskich form językowych, wobec których masz wątpliwości? (zakreśl kółkiem odpowiednią literę)*

   A. nie, nigdy
   B. próbowałam/em, ale coś mnie zniechęciło (napisz co)

   .....

   .....

   C. tak, ale tylko jeśli było to obowiązkowe zadanie domowe
   D. tak, ale tylko okazjonalnie
   E. tak, korzystam z korpusu regularnie.

9. *Jeśli chcesz dodać coś jeszcze na temat korzystania z zadań i materiałów korpusowych, możesz to zrobić tutaj:*

_____

_____

_____

_____

DZIĘKUJĘ! ☺

AL

Survey results – PART I

		Group		Total	Valid percent
		group 1+2	group 3		
N		18	17	35	100
Gender	Female	14	13	27	75
	Male	4	4	8	25
Other studies	No	13	14	27	77.1
	Yes	5	3	8	22.9
Planning MA	No	2	1	3	8.8
	Yes	16	15	31	91.2
English - self assessment (as compared to expected at this point)	Poor	0	1	1	2.9
	Average	6	0	6	17.1
	Good	11	13	24	68.6
	Very good	1	3	4	11.4
Foreign Language before English	German	3	3	6	17.1
	French	0	0	0	0
	Russian	1	0	1	0.3
	Spanish	0	0	0	0
	Latin	0	1	1	0.3
	Others	0	0	0	0
No. of L3s	1	12	6	18	51.4
	2	6	10	16	45.7
	3	0	1	1	2.9
L3 level (highest if more than one)	Elementary	7	6	13	36.1
	Lower-intermediate	5	6	11	31.4
	Intermediate	6	5	11	31.4
	Upper-intermediate	0	0	0	0
	Advanced	0	0	0	0

Detailed survey results not included in the chapter

PART II

Question 4 – Evaluation of task effectiveness

	Task type	Group									
		group 1+2					group 3				
		none	limited	average	high	very high	none	limited	average	high	very high
CORPUS-BASED	NS concordance	0	0	8	8	0	0	0	5	9	2
	parallel corpus	0	2	9	6	1	0	1	3	9	3
	partial transl - conc	0	1	6	10	0	0	0	7	5	3
	EC - conc	0	2	4	8	0	0	1	3	9	3
	gapped concordances	1	0	4	11	0	0	0	5	8	2
	frequency lists	0	6	4	5	1	0	3	6	6	1
SENTENCE-BASED	teacher's examples	0	2	1	13	2	0	0	2	14	1
	translation - sentences	0	3	4	9	2	1	0	7	7	2
	EC - sentences	0	0	6	11	0	1	1	6	3	4
	synthesis	0	2	5	9	2	0	2	6	6	2
	transformations	0	0	5	6	7	0	1	2	7	7
	sentence building	0	1	3	8	6	0	0	4	10	3
TEXT-BASED	translation - fragments	0	2	6	5	4	0	1	5	8	3
	EC - text	0	1	7	8	2	1	0	4	8	3
	text building	1	1	5	6	4	0	0	6	7	3
	word maps	0	1	10	2	3	0	0	8	4	0

Other questions:

		Group		Total	Valid percent
		group 1+2	group 3		
Q2 – No of corpus lesson language problems remembered correctly	0	1	3	4	11.1
	1	7	3	10	28.6
	2	6	8	14	40.0
	3	4	3	7	20.0
Q7 – Attitude	First not OK, no change	3	2	5	14.7
	First OK, then changed	2	0	2	5.9
	First not OK, then changed	10	6	16	47.1
	First OK, no change	3	8	11	32.4
Q9 – Use of corpus in the future as teacher	No	5	1	6	17.1
	Only in advanced or specialized classes (ESP, EAP, one-to-one, etc.)	6	5	11	31.4
	Yes	7	11	18	51.4
Q10 – Own use of corpus	Never	9	4	13	37.1
	Tried but discouraged	0	2	2	5.7
	Yes, only as homework	2	3	5	14.3
	Yes, occasionally	3	8	11	31.4
	Yes, regularly	4	0	4	11.4

Summary in Polish

Korpusy językowe stanowią obecnie jedno z podstawowych narzędzi badań nad językiem. Korzysta się z nich również na wiele sposobów w nauczaniu języków obcych. Redagowanie słowników, podręczników kursowych oraz materiałów pomocniczych do nauki gramatyki, słownictwa oraz sprawności językowych bez korzystania z korpusu językowego jest dziś rzadko spotykane. Większość zastosowań korpusów jest jednak w tej dziedzinie pośrednia: autorzy na podstawie uzyskanych z nich informacji podejmują decyzje dotyczące zawartości owych materiałów, często też czerpiąc z nich autentyczne przykłady użycia danej formy. Dzięki temu materiały te mogą być lepiej dostosowane do potrzeb ucznia, oferując dane językowe odpowiadające faktycznemu użyciu języka w społeczności jego rodzimych użytkowników. Rzadkością jest prezentowanie w podręcznikach danych w formie konkordancji, a informacje statystyczne, choć często podawane w hasłach większych słowników, są bardzo ogólne. Skutkiem tego nauczyciele i uczniowie najczęściej nie zdają sobie sprawy z istnienia korpusów językowych i z tego, jak dalece ich praca i nauka są przez korpusy wspomagane.

Zamiarem autorki niniejszej pracy było wykazanie, że dane z korpusów językowych mogą i powinny zaistnieć w klasie językowej w bardziej bezpośredniej, surowej formie. Korpusy pozwalają przygotować materiały dokładnie odpowiadające potrzebom uczniów, pomagając im w opanowywaniu tych elementów języka obcego, które są szczególnie trudne, na przykład ze względu na interferencję cech języka rodzimego. Materiały te można też dostosować do szczególnych zainteresowań uczniów lub rejestru języka, który odpowiada ich potrzebom zawodowym. Stanowią one źródło autentycznego materiału językowego, najczęściej udostępnianego w postaci konkordancji, które

dzięki swemu unikalnemu formatowi silnie oddziałują na percepcję wizualną i przez to ułatwiają świadome postrzeganie zjawisk i reguł językowych.

Skupienie uwagi na eliminowaniu błędów interferencyjnych było skutkiem doświadczeń autorki związanych z nauczaniem. W bezpośrednim kontakcie z polskimi zaawansowanymi uczniami języka angielskiego i ich problemami w opanowaniu niektórych aspektów gramatyki i leksyki angielskiej, powstała potrzeba znalezienia nowych, bardziej skutecznych sposobów pracy nad doskonaleniem znajomości języka. Powtarzalność, by nie powiedzieć uporczywość pewnych błędów i ich szczególne cechy sugerujące powiązanie z językiem polskim, wskazywały na interferencję jako jedno z głównych źródeł problemów. Fakt, że występowały szczególnie często w zadaniach tłumaczeniowych dodatkowo potwierdzał te przypuszczenia. Materiał językowy zebrany z blogów prowadzonych przez studentów został zebrany w korpus uczniowski i posłużył do szczegółowej do analizy ich użycia języka angielskiego, zarówno pod względem błędów interferencyjnych jak i zjawisk unikania pewnych form lub nadużywania innych. Dzięki temu korpusy pełniły w niniejszej pracy rolę nie tylko źródła danych językowych wykorzystywanych do celów dydaktycznych, ale dostarczały też materiału potrzebnego do zdiagnozowania problemów uczniów zaawansowanych i odpowiedniego skonstruowania badań.

Rozdział pierwszy pracy omawia zagadnienia związane z pojęciem błędu językowego w ogóle oraz błędu interferencyjnego i zjawiska transferu w szczególności. Za zasadnicze kryterium przy definiowaniu pojęcia błędu językowego uznano akceptowalność, to znaczy zgodność danej formy z wewnętrznymi regułami gramatycznymi języka (gramatyczność) i przy tym taki dobór środków językowych, aby brzmiały one naturalnie w swoim kontekście (sytuacyjnym i/lub językowym). Poprawność językowa rozumiana jako zgodność z preskryptywnymi normami językowymi została odrzucona, gdyż jest kryterium zbyt arbitralnym i nieprzystającym do potrzeb współczesnego ucznia zaawansowanego. Uczeń taki musi wykazywać świadomość zróżnicowania cech języka w zależności od rodzaju sytuacji, w której jest on używany i umiejętność dobierania odpowiednich form do kontekstu sytuacyjnego i rejestru.

Sporą część rozdziału stanowi chronologiczny przegląd poszczególnych teorii i podejść do nauczania języka obcego pod względem ich stanowiska wobec błędów językowych. Poczynając od behawioryzmu, poprzez idee Chomsky'ego, Cordera, Selinkera i Krashena, po model kognitywny zaproponowany przez Susan Gass oraz językoznaw-

stwo korpusowe, pojęcie błędu nabierało różnych znaczeń. Dla behawiorystów był to na przykład zły nawyk, Corder widział je jako kluczowe przy testowaniu hipotez, podczas gdy dla zwolenników nauczania komunikatywnego jest to zjawisko marginalne, które można ignorować, o ile znacząco nie zakłóca wymiany informacji pomiędzy nadawcą i odbiorcą.

Kolejny duży fragment rozdziału pierwszego poświęcony jest zjawisku transferu, wielorakim czynnikom za niego odpowiedzialnym, oraz jego wpływowi na proces uczenia się języka, ze szczególnym uwypukleniem współzależności pomiędzy interferencją a fosylizacją. Rozdział kończy się przeglądem klasyfikacji błędów językowych, z podziałem na dwie zasadnicze grupy: taksonomie opisowe i wyjaśniające. Te pierwsze działają powierzchownie, dzieląc błędy na kategorie bez wnikania w ich genezę. Na przykład, powszechnie stosowana taksonomia językowa dzieli błędy w zależności od kategorii językowych, do których należą naruszone przez błąd formy językowe; taksonomia struktur powierzchniowych opisuje sposób, w jaki forma użyta przez ucznia odbiega od formy docelowej. Z kolei druga grupa klasyfikuje błędy próbując znaleźć ich przyczynę: czynniki między- lub wewnątrzjęzykowe, strategie wybrane przez ucznia, techniki zastosowane przez nauczyciela, nieumiejętne korzystanie z materiałów pomocniczych, itp.

W drugim rozdziale pracy w centrum uwagi znajdują się teoretyczne podstawy uczenia się za pomocą danych językowych (DDL). Omawiane jest językoznawstwo korpusowe, jego główne koncepcje (np. zasada otwartego wyboru i zasada idiomu, pojęcie leksykogramatyki) i unikalne techniki analizy języka. Następnie analizowane są procesy wspomagające uczenie się za pomocą danych językowych. Teoria prototypu dotycząca formowania pojęć, podważa klasyczny sposób kategoryzacji oparty na stwierdzeniu obecności pewnego zespołu cech i twierdzi, że kategorie i pojęcia do nich się odnoszące formowane są wokół ich najbardziej typowych przedstawicieli. Dane uzyskane z korpusów językowych uwidaczniają cechy główne i poboczne użycia danej formy językowej, przez co ułatwiają jej klasyfikację, przyswojenie i właściwe stosowanie. Z kolei teoria prymowania leksykalnego postuluje, iż sekwencje współwystępujących ze sobą wyrazów, które zapamiętujemy ucząc się języka, stanowią silnie zespolone bloki. Przy konstruowaniu zdania użytkownik języka sięga przede wszystkim po nie, a dopiero w drugiej kolejności szuka dostępnych reguł składniowych, aby frazy te połączyć ze sobą w spójną, poprawną gramatycznie całość. Struktury gramatyczne są więc

drugorzędne przy formowaniu wypowiedzi, a pierwszorzędną rolę pełnią frazy, kolokacje i koligacje. Ponieważ zgodnie z teorią prymowania leksykalnego uczymy się słów w kontekstach i zapamiętujemy, z jakimi słowami współwystępują, konkordancje wydają się właściwym materiałem do wspomagania tego procesu.

Omawiając językoznawstwo korpusowe nie można nie wspomnieć o gramatyce probabilistycznej, czyli opartej na danych frekwencyjnych pochodzących z korpusów językowych. Nie ma ona bezpośredniego związku z dydaktyką języków obcych, znajduje jednak powszechne zastosowanie w przetwarzaniu języka naturalnego (NLP – *natural language processing*), na przykład w translatorach oraz w aplikacjach, które służą do komunikacji pomiędzy użytkownikiem a komputerem lub stosowane są do automatycznej obsługi klienta.

Bardziej związana z dydaktyką i DDL jest teoria egzemplarzowa języka (*exemplar theory*), według której użytkownik buduje swoją gramatykę i leksykon w oparciu o zapamiętane konkretne wystąpienia zdań i konstrukcji, czyli egzemplarze. Co więcej, każdy napotkany element języka poddawany jest przez użytkownika kategoryzacji. Skutkiem tego każdorazowo dana kategoria może ulec drobnym przeobrażeniom, poprzez wpływ dokonanych za każdym razem rozstrzygnięć co do przynależności danego elementu bądź jego wykluczeniu. Procesy te mogą być świadomie wykorzystane przy nauczaniu języka obcego na poziomie zaawansowanym, gdzie uczeń otrzymuje coraz bardziej skomplikowane dane językowe i musi zrewidować kategorie, którymi dotąd operował, aby dostosować je do nowo poznanych faktów. Prezentacja takich danych za pomocą materiałów korpusowych ułatwia porządkowanie nowych informacji i konstruowanie bardziej precyzyjnych kategorii, reguł i pojęć.

Procesy zachodzące podczas uczenia się wspomaganego technikami DDL mogą być wyjaśnione poprzez różne teorie uczenia się i te właśnie zagadnienia stanowią temat kolejnej części rozdziału drugiego. Autorzy DDL najczęściej powołują się na konstruktywizm kognitywny (lub poznawczo-rozwojowy) Jeana Piageta, aby wyjaśnić jak dostęp do uporządkowanych danych językowych pozwala uczniowi nie tylko rozwinąć większą kompetencję językową, ale i kształtować jego bardziej autonomiczną postawę w procesie uczenia się. Według teorii konstruktywistycznej, wiedza nie jest wprost przekazywana i bezpośrednio przyswajana przez ucznia, lecz musi on ją na swój użytek zrekonstruować i zbudować własne struktury myślowe, aby móc z niej później aktywnie korzystać. Dlatego uczenie się jest bardziej efektywne tam, gdzie zamiast gotowych

odpowiedzi uczeń otrzymuje dane i fakty, które następnie analizuje, obserwując powtarzalne wzorce i dokonując uogólnień. Na tym właśnie polega uczenie się za pomocą danych językowych, przy użyciu materiałów pochodzących z korpusów. Ważna jest tu również otwarta postawa wobec popełnianych błędów, zarówno w produkcji językowej jak też w samej analizie. Są one traktowane jako niezbędny element procesu uczenia się lub wręcz jego początek, gdyż wywołują u ucznia pożądany stan dysonansu poznawczego, który motywuje go do zrekonstruowania swojej wiedzy. Błąd jest więc zarazem wskaźnikiem postępu oraz wyznacznikiem kierunku działania dla nauczyciela i dla ucznia. Procesy uczeniowe, które najczęściej wymienia się w związku z DDL to rozumowanie indukcyjne oraz uczenie się przez odkrywanie. Ich wielką wartością jest to, że angażują one ucznia intelektualnie, prowokując do samodzielnego wysiłku zamiast podsuwać gotowe rozwiązania. Ten model kształcenia jest obecnie uznawany w pedagogice za właściwy i bardziej adekwatny do potrzeb współczesnego świata.

Innym spojrzeniem na procesy uczeniowe zaangażowane w DDL jest koneksjonizm, lub szerzej teoria uczenia się z punktu widzenia neurofizjologii i teorii sieci neuronowych. W takim ujęciu uczenie się z wielu podanych przykładów nie polega na formowaniu abstrakcyjnych reguł, a jedynie jest wzmacnianiem połączeń neuronowych odpowiadających danej konfiguracji elementów językowych, co skutkuje ich utrwaleniem i łatwiejszym przywołaniem w sytuacji, gdy zaistnieją podobne warunki do ich użycia. Kluczowym elementem tej teorii jest naturalna u ludzi skłonność do rozpoznawania wzorców w przedstawionym materiale i utrwalania ich w pamięci. Koneksjonizm przeczy więc istnieniu reguł językowych, jako konstruktów biorących czynny udział w formowaniu zdań, widząc je raczej jako próby opisu tego, co w języku zachodzi niejako samoistnie, w oparciu o procesy neurofizjologiczne. Badania nad sieciami neuronowymi są dalekie od „żywej" dydaktyki, gdyż eksperymenty takie przeprowadzane są jedynie na specjalnie skonstruowanych, wyidealizowanych modelach tychże sieci. Także ilość i jakość danych, które modele te są w stanie przetworzyć jest minimalna, w porównaniu z naturalnym językiem i jego złożonością. Niemniej modele te otwierają drogę do zrozumienia procesów uczenia się od strony fizjologicznej, a analizy ich działania potwierdzają przydatność uczenia się z wielokrotnych przykładów użycia tego samego elementu językowego, czyli tego, co oferuje w swych materiałach DDL.

W rozdziale omawiane są też teorie, które skupiają się na akwizycji języka, a nie na uczeniu się w sensie ogólnym. Pierwsza z nich to podejście naturalne Krashena,

umieszczone tu ze względu na szczególną rolę, jaką przypisywało ono właściwemu doborowi wejściowych danych językowych udostępnianych uczniowi języka obcego. Druga to podejście leksykalne Michaela Lewisa, omawiane z powodu wyjątkowej roli, jaką przypisuje on zbitkom wyrazowym i kolokacjom w procesie uczenia się języka i osiąganiu naturalności w jego użyciu, a także ze względu na przypisaniu gramatyce roli receptywnej sprawności językowej, która pozwala uczniowi rozpoznawać różnice znaczeniowe i stylistyczne wynikające z poszczególnych wyborów gramatycznych. Tak jak w DDL, priorytetem jest tu naturalność i dobór środków językowych odpowiedni do kontekstu, w jakim one występują.

Ostatnią z omawianych teorii jest nauczanie zorientowane na formę, (*form-focused instruction*, FFI), które należy postrzegać bardziej jako pewną tendencję, czy też trend w dydaktyce języków obcych, a nie spójną, w pełni uformowaną teorię. DDL niewątpliwie mieści się w tej właśnie tradycji i dlatego przegląd tez i propozycji mieszczących się w jej ramach uznany został za niezbędny w niniejszej dysertacji. W ramach FFI, ogólnie przyjęty jest podział na zorientowanie na *formy* (działania, których głównym celem jest doskonalenie właściwego doboru i budowy wybranych przez nauczyciela, z góry określonych form językowych) oraz zorientowanie na *formę* (kierowanie uwagi uczniów na poprawne użycie tych elementów języka, które wyłaniają się podczas działań nakierowanych na komunikację i skupionych przede wszystkim na przekazywanym w trakcie wykonywania zadań znaczeniu). Podział ten plasuje DDL w grupie pierwszej. Faktem jest, że lekcje oparte na materiałach korpusowych przeważnie ukierunkowane są na uprzednio zaobserwowane przez nauczyciela problemy uczniów, jednak charakter tych materiałów, szczególnie w wersji drukowanej, nie pozwala na spontaniczne ich użycie. Są to więc lekcje ściśle zaplanowane, prowadzone z nieskrywaną intencją doskonalenia poprawności językowej uczniów i podnoszenia ich świadomości językowej. Głównym podłożem teoretycznym dla FFI jest teoria świadomego spostrzegania Richarda Schmidta (1990), według której dany element języka nie wejdzie do repertuaru językowego ucznia, dopóki nie zostanie przez niego *świadomie* spostrzeżony w przedstawionym mu materiale. Nie wystarczy więc tylko powierzchowne odnotowanie wystąpienia danej formy, lecz konieczne jest poświęcenie jej uwagi i objęcie jej świadomą refleksją. Wtedy możliwe jest przeniesienie danego elementu z pamięci krótkotrwałej do długotrwałej, co z kolei konieczne jest do skutecznego nauczenia się go. Inne rozróżnienia często stosowane do opisu działań w klasie językowej zaklasyfikowa-

łyby nauczanie za pomocą DDL jako eksplicytne (otwarcie kierujące uwagę uczniów na formę językową), indukcyjne (prowadzące od przykładów do sformułowania reguły), z uporządkowanymi danymi (w odróżnieniu od ćwiczenia poprzez produkcję) oraz proaktywne (nakierowane na zapobieganie dalszym błędom w stosowaniu danej formy). Omówienie nauczania zorientowanego na formę kończy się przeglądem cech charakterystycznych uczniów, których potrzeby wskazują na największą celowość wyboru takiego właśnie podejścia do nauki języka. Uczeń taki jest na ogół dorosły, zaawansowany w nauce języka, dobrze wykształcony, często operuje językiem pisanym w rejestrze formalnym, dla celów zawodowych. Te same zapewne cechy będą dobrym wskazaniem do stosowania technik DDL.

Rozdział trzeci dysertacji przybliża czytelnikowi ogólne właściwości nauczania za pomocą danych językowych, prezentuje jego główne techniki oraz wcześniej publikowane badania nad ich skutecznością. Na początku przedstawiona jest krótka geneza samego pojęcia DDL oraz przegląd pierwszych materiałów do nauki języka opracowanych w oparciu o korpusy językowe. Inicjatorem podejścia zwanego nauczaniem opartym na danych językowych oraz autorem samego terminu DDL jest Tim Johns, który swoje nowatorskie rozwiązania zaprezentował we wczesnych latach 90-tych. Nauczanie takie łączy dwie podstawowe cechy: wykorzystanie danych z korpusów językowych oraz uczenie się przez odkrywanie. Połączenie tych właściwości pozwala na zmianę tradycyjnych ról w klasie językowej: uczeń i nauczyciel stają się do pewnego stopnia partnerami nie tylko w poszukiwaniu właściwych odpowiedzi, ale również w stawianiu pytań. Uczniowie wspólnie z nauczycielem analizują konkordancje lub dane frekwencyjne, aby dojść do reguł językowych i pewnych uogólnień odnośnie stosowania poszczególnych form. W nauczaniu takim niezmiernie ważna jest otwarta postawa nauczyciela, gdyż obserwacje poczynione przez uczniów mogą różnić się od tych oczekiwanych przez nauczyciela, ale mogą okazać się równie wartościowe. W DDL nauczyciel nie jest „wyrocznią", która rozsądza, co jest właściwe i poprawne a co nie. Staje się nauczycielem poszukującym. Od ucznia techniki te wymagają z kolei aktywności intelektualnej, a szczególnie rozumowania indukcyjnego, które konieczne jest do rozpoznawania wzorców w dużej ilości udostępnionych danych. Nauczyciel musi być dobrym obserwatorem i w razie potrzeby umieć nakierować uczącego się na właściwe tory rozumowania, nie podając jednocześnie gotowej odpowiedzi. Ta zmiana ról wydaje się dobrze dostosowana do oczekiwań i postaw dzisiejszych uczniów, zwłaszcza za-

awansowanych, którzy są przeważnie bardziej dojrzali i przez to gotowi do przejęcia części inicjatywy w procesie nauki języka.

DDL może być postrzegane jako kompromis pomiędzy skrajnie przeciwnymi trendami w nauczaniu języka, a w szczególności gramatyki. Z jednej strony jest więc postrzeganie gramatyki jako produktu, ze ściśle wyznaczonymi, sztywnymi regułami, a słownictwem w roli drugorzędnej. Po stronie przeciwnej znajduje się gramatyka jako proces, czyli wiedza proceduralna, niezbędna do konstruowania zdań i ich analizowania w celu osiągnięcia komunikacji z innymi użytkownikami języka. Tutaj większy jest nacisk na precyzyjne i stosowne do zamierzonej funkcji słownictwo, a priorytetem jest przekazanie informacji. DDL znajduje się pomiędzy tymi dwiema skrajnościami, szczególnie ze względu na równoważne traktowanie gramatyki i słownictwa. Językoznawstwo korpusowe, a za nim DDL, postuluje zniesienie rozgraniczenia pomiędzy tymi dwiema zazębiającymi się ściśle kategoriami i stworzenie w ich miejsce jednej: leksykogramatyki.

DDL charakteryzuje się pewnymi unikalnymi cechami, które wniosły nową jakość do dydaktyki języków obcych. Oprócz właściwości omówionych powyżej, warto wspomnieć o nowej roli, jaką przypisuje ono komputerowi w procesie uczenia się języka. Przed pojawieniem się DDL główną funkcją komputera było zastępowanie nauczyciela. Tak konstruowane były i często są nadal programy komputerowe do nauki języków obcych instalowane z nośników optycznych (CD-ROM lub DVD) lub oferowane w Internecie: komputer oferuje zestaw zadań do rozwiązania i następnie informuje użytkownika o uzyskanym wyniku i ewentualnie o popełnionych błędach. W nauczaniu za pomocą danych językowych komputer pełni zupełnie inną funkcję; staje się źródłem danych, które uczeń poddaje analizie i samodzielnie dochodzi do uogólnień. Zmiana ta ma charakter nie tylko techniczny, ale przede wszystkim dydaktyczny. Zmienia ona rolę ucznia z klienta lub odbiorcy „usługi edukacyjnej" na rolę badacza lub odkrywcy. Komputer staje się narzędziem w rękach ucznia zamiast dyktować mu przebieg i treść jego nauki. Otwiera to uczniowi nowe możliwości, przede wszystkim oferując mu większą autonomię w procesie uczenia się i motywując do stawiania własnych pytań i poszukiwania nań odpowiedzi. Tego typu postawa zalecana jest przez *CEFR* (Europejski System Opisu Kształcenia Językowego), gdyż pozwala uczniowi rozwijać swoje umiejętności językowe długo po ukończeniu formalnej edukacji w tym zakresie. Inna cecha wyróżniająca DDL to elastyczność: nauczyciel może przygotować materiały do-

stosowane do unikalnych potrzeb swoich uczniów, które wcześniej zbadał lub zaobserwował w swej codziennej praktyce.

Po charakterystyce DDL następuje przegląd technik nauczania, które zwolennicy tego podejścia stosują i polecają. Techniki te są tu uporządkowane według typów korpusów, z których pochodzą materiały wykorzystywane do ich przygotowania: konkordancje z korpusów tekstów pochodzących od rodzimych użytkowników języka docelowego, z korpusu języka pierwszego, korpusu równoległego, korpusu uczniowskiego, materiały niewykorzystujące konkordancji, materiały do indywidualnej pracy z korpusem oraz komputerowe aplikacje do nauki języka oparte na danych korpusowych. Przykładowe zadania to analiza porównawcza, wypełnianie luk, częściowe tłumaczenie oraz rozpoznawanie i poprawianie błędów.

Kolejna sekcja rozdziału trzeciego zawiera skrótowy przegląd badań nad skutecznością nauczania zorientowanego na formę (FFI) oraz bardziej szczegółową analizę badań nad skutecznością technik DDL. Na podstawie części pierwszej sformułować można ogólny wniosek, że najbardziej skuteczne jest nauczanie w trybie eksplicytnym, czyli otwarcie kierujące uwagę uczniów na formę językową. Konkluzje po analizie badań nad nauczaniem w oparciu o dane językowe były już mniej jednoznaczne, gdyż wyniki eksperymentów różniły się od siebie. Generalnie można powiedzieć, że w większości eksperymentów DDL wykazywało nieznaczną przewagę nad technikami konwencjonalnymi, choć była ona na ogół statystycznie nieistotna. Rozdział kończy się rozważaniami na temat wpływu stylu uczenia się na efekty, jakie uczeń może uzyskać za pomocą technik DDL, oraz krótkim omówieniem ich negatywnych aspektów, które zostały zaobserwowane przez niektórych badaczy i samą autorkę.

Ostatni rozdział pracy relacjonuje przebieg badań przeprowadzonych dla potrzeb niniejszej dysertacji oraz ich wyniki. Ich celem było zbadanie skuteczności nauczania za pomocą danych językowych i porównanie ich pod tym względem do technik bardziej konwencjonalnych. Hipoteza badawcza postawiona na początku badań postulowała, że techniki DDL są bardziej skuteczne w eliminowaniu błędów interferencyjnych u uczniów zaawansowanych niż techniki konwencjonalne. Oprócz głównej hipotezy, praca stawia trzy pytania badawcze. Pierwsze z nich dotyczy rodzaju błędów interferencyjnych występujących najczęściej u polskich zaawansowanych uczniów języka angielskiego; przedmiotem drugiego pytania jest istnienie współzależności pomiędzy skutecznością technik korpusowych a indywidualnymi cechami uczniów, zaś trzecie

pytanie badawcze odnosi się do oceny technik DDL i ich skuteczności przez samych studentów biorących udział w badaniu. Projekt badawczy składał się z trzech etapów: analizy błędów i produkcji językowej uczniów zaawansowanych, badania eksperymentalnego, oraz ankiety. Każdy kolejny etap opierał się na wynikach poprzedniego, co stworzyło obszerny zasób danych, budując jednocześnie dość spójny obraz problemu. W ramach etapu pierwszego przeanalizowano pisemne prace egzaminacyjne studentów trzeciego roku filologii angielskiej w KJO, aby zdiagnozować problemy interferencyjne, z jakimi się zmagają. Wyniki tej analizy posłużyły następnie do komputerowej analizy błędów, przeprowadzonej na korpusie uczniowskim (ponad 200 tys. słów) zbudowanym z tekstów zamieszczonych przez studentów kilku roczników tej samej populacji na blogach prowadzonych w języku angielskim w ramach zajęć w KJO. Celem pierwszego etapu badań była selekcja takich zagadnień językowych, które faktycznie stanowiły trudność dla ich uczestników i mogły stać się przedmiotem lekcji eksperymentalnych w etapie kolejnym.

Procedura badawcza głównej fazy badań polegała na przeprowadzeniu lekcji dotyczących konkretnych zagadnień językowych, które poprzez analizę błędów zostały uznane za częste źródła pomyłek i w związku z tym wymagały interwencji dydaktycznej. Celem lekcji było podniesienie świadomości językowej studentów, co do wybranych zagadnień i zminimalizowanie niepoprawnego stosowania omawianych form językowych. Lekcje były przeprowadzane w dwóch grupach, których role się zmieniały: status grupy eksperymentalnej i kontrolnej był im przypisywany naprzemiennie, tak, aby wszyscy uczestnicy badań mogli doświadczyć nauczania technikami DDL i wyrobić sobie opinię na ich temat. Lekcje zestawiono więc w trzy podwójne moduły, z których każdy dotyczył innego obszaru leksykogramatyki (składnia, wybór leksykalny, policzalność). W ramach modułu jedna grupa studentów odbywała lekcję opartą o dane korpusowe, pełniąc wtedy rolę grupy eksperymentalnej, oraz lekcję konwencjonalną, pełniąc rolę grupy kontrolnej. W grupie równoległej stosowano taki sam układ, przy czym materiał lekcji DDL poprzedniej grupy był tutaj przedstawiany konwencjonalnie i na odwrót. W ten sposób uzyskano dwa zestawy porównywalnych danych, na podstawie których możliwe było poczynienie szerszych obserwacji. Pomiar skuteczności lekcji opierał się na danych uzyskanych z pretestu, posttestu i testu opóźnionego (technika tłumaczeniowa): zmienną zależną była tutaj wartość oznaczona jako przyrost (*gain*), to znaczy różnica pomiędzy ilością błędów przed interwencją dydaktyczną i po niej, a tak-

że po upływie czterech tygodni. W kontekście niniejszych badań przyrost należy rozumieć jako zwiększenie poprawności, czyli zmniejszenie ilości błędów. Dane dotyczące każdej lekcji poddane zostały testom statystycznym w celu porównania dwóch prób niezależnych (wyników po lekcji korpusowej i konwencjonalnej) i uzyskania potwierdzenia lub odrzucenia głównej hipotezy badawczej, a więc ustalenia czy nauczanie technikami DDL jest bardziej skuteczne od konwencjonalnego, czy też nie. Ponieważ w większości przypadków stwierdzono za pomocą testu Shapiro-Wilka, że rozkład danych odbiegał od normalnego, zastosowano tam testy nieparametryczne, to jest test U Manna Whitneya. W kilku przypadkach dodatkowo przeprowadzono test Wilcoxona (nieparametryczny test dla dwóch grup zależnych), aby stwierdzić, czy lekcje były w ogóle skuteczne. Warto dodać, że testy te dały wynik pozytywny.

W trzeciej fazie badań, aby uzupełnić obraz sytuacji uzyskany w wyniku badań ilościowych i dokonać triangulacji, zastosowano również analizę jakościową w postaci ankiety. Za jej pośrednictwem studenci oceniali skuteczność rożnych typów zadań językowych (DDL i konwencjonalnych) i wyrażali swoje opinie na temat zadań i narzędzi korpusowych. Ankieta posłużyła też do zebrania dodatkowych informacji o uczestnikach, potrzebnych do analizy korelacji.

Wyniki przeprowadzonych badań są dość niejednoznaczne. Analiza błędów wykazała liczne problemy o podłożu interferencyjnym, lecz ich zróżnicowanie u studentów było ogromne i wyłonienie kilku zagadnień językowych stanowiących wyzwanie dla dużej większości uczestników badań okazało się dość trudne. Sama analiza błędów za pomocą korpusu uczniowskiego i narzędzi do analizy tekstu dostarczyła jednak sporo informacji o interjęzyku polskich zaawansowanych uczniów języka angielskiego, ujawniając przykłady użycia strategii uniku o podłożu interferencyjnym i innych zjawisk o podobnym charakterze. Najczęstszymi typami błędów interferencyjnych okazały się: nierozróżnienie (*underdifferentiation*, czyli użycie formy, która ma ekwiwalent w języku ojczystym ucznia wspólny z inną, w danym kontekście właściwszą formą), kalka językowa, oraz błędy wynikające z braku danej kategorii w języku ojczystym (np. opuszczanie przedimków).

Hipoteza badawcza postawiona na początku pracy nie została potwierdzona. Według wyników przeprowadzonych badań, skuteczność technik DDL nie jest większa od tradycyjnie stosowanych technik nauczania, choć w przypadku zagadnień językowych o charakterze leksykalnym rezultaty wykazały nieznaczną przewagę technik kor-

pusowych. Analiza nie wykazała jednak, aby różnice te były istotne statystycznie. Co więcej, skuteczność DDL okazała się bardziej niestabilna, z dość dużymi różnicami pomiędzy poszczególnymi lekcjami w danej grupie. Trudno stwierdzić, co dokładnie mogło wywołać takie wahania, ale mogą one być związane z naturą materiału językowego, który był prezentowany i przyswajany w czasie lekcji. Tego zjawiska nie zaobserwowano w przypadku technik konwencjonalnych.

Wyniki ankiety były bardziej zachęcające. Zdecydowana większość studentów oceniła lekcje wykorzystujące dane korpusowe pozytywnie, zwracając uwagę na ich innowacyjny charakter oraz fakt, że zmuszają one uczniów do samodzielnego myślenia. Interesujące były też wypowiedzi dotyczące słabych stron DDL. Uczestnicy wymieniali tu często przytłaczającą ilość informacji oraz mało interakcyjny charakter tego typu działań w klasie. Niektórzy studenci wyrażali uznanie dla technik DDL i ich skuteczności, zastrzegając jednak przy tym, że im osobiście one nie odpowiadają i są dla nich nieatrakcyjne. Znalazło się też jednak kilkoro entuzjastów DDL, którzy na skutek odbytych w ramach badań zajęć zainteresowali się korpusami językowymi i stali się ich regularnymi użytkownikami.

Badanie korelacji nie przyniosło jednoznacznych odpowiedzi. Nie stwierdzono zależności pomiędzy aprobatą dla technik korpusowych a ich skutecznością dla poszczególnych uczniów. Także poziom znajomości języka mierzony przez standaryzowane testy językowe nie zdaje się mieć wpływu na sukces nauczania za pomocą danych językowych. Stwierdzono jedynie zależność pomiędzy wynikami testów przeprowadzanych w ramach kursu gramatyki a jedną zmienną: wynikami przyrostu dla jedynej lekcji opartej na danych korpusowych przeprowadzonej w obydwu grupach właśnie w celu uzyskania jednolitych danych dla całej próby. Wynik ten wskazuje, że niewielki rozmiar grupy badawczej mógł mieć wpływ na niskie wskaźniki korelacji w przypadku danych dla pozostałych lekcji. Potrzebne byłyby dalsze badania, aby ustalić, czy rezultaty te były właściwe, czy też większa liczebność grupy mogłaby wykazać pewne współzależności pomiędzy cechami indywidualnymi uczniów a skutecznością DDL w eliminowaniu błędów interferencyjnych.

Pomimo niestwierdzenia istotnych różnic ilościowych, wyniki ankiety oraz sama charakterystyka technik DDL zawarta we wcześniejszych rozdziałach sugerują, iż warto je stosować z uczniami na poziomie zaawansowanym. Mają one bowiem wiele zalet innego rodzaju: oferują uczniom znaczne korzyści w zakresie uczenia się przez całe

życie (*lifelong learning*), którego znaczenie podkreślane jest w rekomendacjach dotyczących nauczania języków obcych opracowanych przez Radę Europy. Korpus to narzędzie wspomagające samodzielny rozwój ucznia, jego autonomię i postawę partnerską w relacji nauczyciel/uczeń. Praca z danymi korpusowymi to jeden ze sposobów przygotowania ucznia do dalszego doskonalenia znajomości języka po ukończeniu edukacji formalnej. Ma to szczególne znaczenie dla studentów filologii, którzy z tego typu narzędzi mogą z powodzeniem korzystać w przyszłej pracy zawodowej, nie tylko w dydaktyce, ale również w tłumaczeniach, badaniach naukowych i innych zadaniach wymagających szczególnej dbałości o właściwy do funkcji i kontekstu dobór środków językowych.

Pewne obiektywne ograniczenia, szczególnie co do liczby uczestników oraz ilości czasu, jaki mógł być wykorzystany na lekcje eksperymentalne, miały wpływ na wyniki przedstawionych tu badań. Aby uzyskać szerszą wiedzę na temat skuteczności nauczania za pomocą danych językowych należałoby powiększyć grupę badawczą oraz przeprowadzić więcej lekcji eksperymentalnych i kontrolnych. Pozostaje nadzieją autorki niniejszej pracy, że temat ten będzie nadal badany, a nowe dane ukażą pełniejszy i być może bardziej pozytywny obraz technik DDL, przedstawiając ich potencjał w sposób w pełni satysfakcjonujący.